EUROPEAN JOURNAL OF CANCER 47 (2011) 854–863

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejconline.com

Randomised controlled trials in oncology closed early for benefit: Trends in methodology, results, and interpretation

Christopher M. Booth *, *Pavlo Ohorodnyk, Liting Zhu, Dongsheng Tu, Ralph M. Meyer*

NCIC Clinical Trials Group and Queen's University Cancer Research Institute, Kingston, Canada K7L 3N6

ARTICLE INFO

Article history:
Received 23 August 2010
Received in revised form 2 December 2010
Accepted 7 December 2010
Available online 4 February 2011

Keywords:
Clinical trials
Data safety monitoring board
Outcomes research
Interim analysis

ABSTRACT

Purpose: To assess methodology, results and interpretation of oncology randomised controlled trials closed early for benefit (RCTCEB).

Methods: Structured literature search (1950–2008) to identify all published oncology RCTCEB. We then searched for related follow-up articles and conference abstracts to evaluate whether study results and conclusions changed with longer follow-up. A standardised data abstraction process captured information related to statistical methodology, details of interim analyses, results and conclusions. Original articles and follow-up reports were compared for results of primary end-point and author conclusions.

Results: We identified 71 RCTCEB. In 16 articles (23%) the study primary end-point was not explicitly stated. Most trials were open to accrual (47/71, 66%) at the time of closure. Formal interim analysis was performed in 65 (92%) trials of which 72% (47/65) was reported as planned; 82% (53/65) reported stopping rules. Trials on average accrued 75% of the planned sample size. Amongst the 23 (32%) RCTCEB with follow-up reports, in only one case did the study results or conclusions change substantially.

Conclusions: While the majority of oncology RCTCEB follows rigourous methodological principles, an important percentage includes limitations in design and/or analysis. Amongst the 23 studies with subsequent follow-up reports, initial results were confirmed in 22 (96%).

1. Introduction

Early closure of randomised controlled trials (RCTs) based on interim observation of large differences between treatment arms is being increasingly reported. While some suggest that oncology RCTs closed early for benefit (RCTCEB) may overestimate benefits[1–4], a recent analysis of National Cancer Institute (NCI) cooperative group trials observed that treatment effects were maintained with longer follow-up.[5]

Risks of false-positive findings of RCTCEB are well recognised and relate to statistical issues, including the inherent instability of early results and multiple 'looks at the data'. Therefore, rigourous analytical methodologies for interim analyses have been proposed, including the use of stopping rules associated with *p*-value thresholds less than 0.05 and consideration of alpha-spending to preserve an overall significance level of 0.05.[6,7] Although recent overviews of oncology RCTCEB have confirmed that early closure is increasingly common[2,3,5], descriptions of the methodology of these trials are limited. Furthermore, only a single report dedicated to cancer trials has described the long-term follow-up of these trials and this report was restricted to trials conducted by NCI cooperative groups.[5] We, therefore, undertook this study to evaluate statistical methodologies of RCTCEB and to com-

* *Corresponding author:* Address: NCIC Clinical Trials Group and Queen's University Cancer Research Institute, 10 Stuart Street, Kingston, ON, Canada K7L 3N6. Tel.: +1 613 533 6430; fax: +1 613 533 2411.
E-mail address: boothc@kgh.kari.net (C.M. Booth).
0959-8049/$ - see front matter
doi:10.1016/j.ejca.2010.12.005

pare study conclusions at the time of closure with reports of longer follow-up.

2. Methods

2.1. Identification of RCTs closed early for benefit

The study cohort includes published phase three oncology RCTs closed early for benefit (RCTCEB). We included trials where benefit was observed with the stated experimental arm or with an active treatment control arm. We excluded RCTs closed for reasons other than efficacy (i.e. regulatory/administrative decision, results of other studies, poor accrual, futility, toxicity) and articles reporting exploratory, subset or pooled analyses were excluded. RCTs closed for futility were defined as those RCTs closed in which there was no observed difference between the study arms; the experimental arm was worse than the best supportive care/placebo; or a new therapy is added to the standard of care and was inferior to the standard of care alone. Eligible studies were identified from a comprehensive literature search, citations from other published overviews and investigator files.

MEDLINE (1950 – 31st May 2008) and EMBASE (1980 – 31st May 2008) were searched using the key words; RCT, phase III, oncology; early termination (e.g. stopped, halted, closed, premature, early); Data Safety Monitoring Board (DSMB); interim analysis; and unblinding. To increase the sensitivity of our searches, multiple key word combinations were used. Preliminary results were pooled and screened for eligibility. English-language abstracts were reviewed to identify direct or indirect references to premature closure of the study. When an abstract used the term 'interim analysis', but an indication of trial closure and/or the reason for this were not indicated, the full text of the article was reviewed. We also included additional RCTCEB known to investigators from overviews recently published by Wilcox et al.[2], Trotta et al.[3], Korn et al.[5], and our own group[8,9] and RCTs from our personal files.

2.2. Identification of follow-up analyses

To explore whether interim results leading to study closure change substantially with follow-up, we also searched for updated RCT results. Using the names of first and last authors, we searched MEDLINE and EMBASE for articles published up to 31st May 2008 and searched electronically for abstract presentations at the annual meetings of the American Society of Clinical Oncology (1995 – 31st May 2008) and the American Society of Hematology (2004 – 31st May 2008).

2.3. Data abstraction

Data capture forms were developed for the initial RCTCEB publication, follow-up articles, and related abstracts. To guide data abstraction, a manual was created and forms were piloted on a series of 10 trials before undergoing final revision.

Variables captured included details of study design, sponsorship, statistical methodology, results and author conclusions. Author conclusions were assigned a score from 1 to 7 based on a scale developed by Ridker and Torres[10]; 4/7 for a neutral statement, 7/7 for strong endorsement of the experimental arm and 1/7 for strong endorsement of the control arm (see Box 1). As previously defined[8,9], author conclusions were deemed to endorse the experimental arm if the author score was 6 or 7 and to endorse control arm if the author score was 1 or 2. Industry support was determined based on explicit statements in the article and affiliations of study authors. For solid tumour trials, early-stage refers to neoadjuvant/adjuvant therapy and advanced-stage refers to palliative therapy for metastatic disease. For haematologic malignancies, early-stage refers to first-line therapy and advanced-stage refers to trials in the relapsed or refractory settings. We used dates of subject enrollment and pre-specified sample size to determine whether the RCTCEB was closed while study accrual was ongoing.

Box 1 Grading scale for authors' conclusions.

Score (/7)	Authors' conclusion
1	Standard of care (SOC) was significantly better and/or highly preferred
2	The SOC was significantly better and/or moderately preferred
3	The SOC was non-significantly better
4	The SOC and newer treatment were equivalent
5	The newer treatment was non-significantly better
6	The newer treatment was significantly better and/or moderately preferred
7	The newer treatment was significantly better and/or highly preferred

2.4. Comparison of initial results with longer follow-up

Results of the primary end-point and author conclusion score were compared with related follow-up articles and abstracts where available. Author conclusions were considered discordant if there was a change in the recommendation to adopt (or not adopt) the experimental regimen as assessed by a change in score from <5 to $\geqslant 5$ (or vice versa) for superiority trials, or from <4 to $\geqslant 4$ (or vice versa) for non-inferiority.[9]

2.5. Statistical analysis

Descriptive statistics were used to summarise data. To describe effect size in a consistent manner across trials, hazard ratios (HR) for the primary end-point in RCTCEB favouring the control arm were converted to 1/HR. The characteristics of RCTCEB were compared by date of publication (1985–1999 versus 2000–2008), sponsorship (cooperative group versus non-cooperative group; industry versus non-industry), trial arm demonstrating benefit (experimental versus control) and whether a follow-up report was identified (yes versus no).

Proportions between groups were compared using Fisher's exact test; means were compared with the *t*-test. All analyses were performed using SAS 9.1.

3. Results

Results of the literature search are shown in Fig. 1. The combined literature searches yielded 1219 articles; 1133 were not randomised controlled trials with early closure. Of the remaining 86 RCTs, 36 were excluded as closure was because of futility ($n = 23$), toxicity ($n = 3$) or other reasons ($n = 10$). Six other studies were excluded because they were only partially closed ($n = 3$), were not in English ($n = 1$), or represented pooled ($n = 1$) or re-randomisation analyses ($n = 1$). Twenty-seven additional RCTCEB were identified from published overviews of oncology RCTs[2,3,5,8] and investigator files. The final study cohort included 71 RCTCEB; 58 (82%) favouring the experimental arm and 13 (18%) favouring the active treatment control arm.

Study cohort characteristics are shown in Table 1. The most common disease sites were breast (12/71, 17%), lung (9/71, 13%) and haematologic malignancies (20/71, 28%) and most publications were in high-impact medical journals. Sixty-one percent (43/71) of reports was published since 2000, 58% (41/71) was conducted by cooperative groups and 31% (22/71) was supported by industry. The study intervention involved systemic therapy (cytotoxic chemotherapy, hormonal or targeted agent) in 83% (59/71) of trials.

Aspects of study design and statistical methodology are shown in Table 2. Almost all studies (66/71, 93%) were parallel group superiority trials. Median sample size was 312 patients (range 27–13,388). In 16 articles (23%), the study primary endpoint was not explicitly stated. Intention-to-treat (ITT)

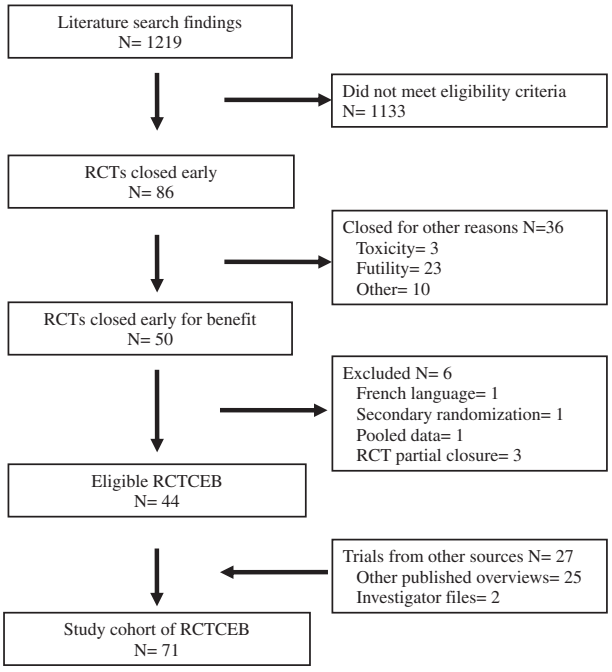

Fig. 1 – Results of search strategy for randomised controlled trials closed early for benefit (RCTCEB).

Table 1 – Study characteristics of RCTs closed early for benefit ($n = 71$)[a].

Disease site	
Breast	12 (17%)
Lung	9 (13%)
Myeloma	6 (8%)
Leukaemia	9 (13%)
Lymphoma	5 (7%)
Other	30 (42%)
Journal	
New Engl J Med	20 (28%)
J Clin Oncol	22 (31%)
Blood	5 (7%)
JNCI	3 (4%)
Lancet	8 (11%)
Other	13 (18%)
Year published	
Pre-1985	0 (0%)
1985–1989	6 (8%)
1990–1994	9 (13%)
1995–1999	13 (18%)
2000–2004	19 (27%)
2005–2008	24 (34%)
Setting[b]	
Early stage	38 (54%)
Advanced	30 (42%)
Supportive care	1 (1%)
Prevention	2 (3%)
Study origin	
North America	39 (55%)
Europe	26 (37%)
Other	6 (8%)
Participants	
National	41 (58%)
International	30 (42%)
Organisation	
Multicenter	63 (89%)
Co-operative group	41 (58%)
CTEP coop group	25 (35%)
Non-CTEP coop group	16 (23%)
Sponsorship	
For-profit	12 (17%)
Non-profit	41 (58%)
Mixed	10 (14%)
Not known	8 (11%)
Intervention	
Chemotherapy	46 (65%)
Placebo/observation	12 (17%)
Hormonal therapy	5 (7%)
Targeted agent	13 (18%)
Complementary medicine	1 (1%)
Surgery	2 (3%)
Radiotherapy	8 (11%)
Other	9 (13%)
Study arm showing benefit	
Experimental	58 (82%)
Control	13 (18%)

[a] Percentages may not add up to 100% due to rounding.

[b] For solid tumour trials early-stage refers to neoadjuvant/adjuvant therapy and advanced-stage refers to palliative therapy for metastatic disease. For haematologic malignancies, early-stage refers to first-line therapy and advanced-stage refers to trials in the relapsed or refractory settings.

Table 2 – Design characteristics of RCTs closed early for benefit (*n* = 71).

Variable	All randomised controlled trials closed early for benefit (RCTCEB) (*n* = 71)[a]
Statement of primary end-point	
Explicit	55 (77%)
Implied	15 (21%)
Not stated	1 (1%)
Nature of primary end-point	
OS	25 (35%)
DFS/RFS/EFS	16 (23%)
TTP/PFS	14 (20%)
RR	8 (11%)
Other	7 (10%)
Missing	1 (1%)
Study design	
Intention to treat all randomised	42 (59%)
Power calculation reported	59 (83%)
Median hazard ratios [HR] (range)	0.69 (0.20, 0.88)
Median power (range)	0.80 (0.80, 0.96)
Median target sample size (range)	354 (44, 4800)
With Data Safety Monitoring Board (DSMB)	44 (62%)
Statement of independence	20 (28%)
At time of early closure	
Accrual ongoing	47 (66%)
Therapy ongoing	62 (87%)
Interim analyses	
Was IA performed?	
Yes	65 (92%)
No	0 (0%)
Not stated	5 (7%)
Continuous sequential	1 (1%)
Amongst trials with interim analysis (n = 65)	
Planned	47/65 (72%)
Unplanned	8/65 (12%)
Not stated	8/65 (12%)
Median number of IA planned (range)	2 (1, 5)
Was IA end-point specified?	
Yes	40/65 (62%)
If yes was it as primary EP?	34/40 (85%)
IA triggered by	
Number of events	20/65 (31%)
Time of study duration in months	5/65 (8%)
Accrual targets	18/65 (28%)
Not stated	24/65 (37%)
Was stopping rule stated?	
Yes	53/65 (82%)
If yes, stopping rule was	
Numeric boundary[b]	20/53 (38%)
O'Brien Fleming	27/53 (51%)
Pocock	2/53 (4%)
Sequential	2/53 (4%)
Other method	1/53 (2%)
p-Value provided for stopping rule	
Yes	30/53 (57%)
Median *p*-value threshold for stopping (range)	0.0080 (0.00025, 0.05)

[a] Denominator is 71 unless otherwise noted. Percentages may not add up to 100% due to rounding.
[b] Numeric boundary refers to trials in which an explicit *p*-value threshold was stated as the stopping rule with no other stopping rule methodology specified.

analysis of all randomised patients was performed in only 59% (42/71). Most trials were open to accrual (47/71, 66%) at the time of early closure. In 24 trials accrual was completed and the trials were in the follow-up phase when the interim data led to early 'closure' and release of study results. Formal interim analyses were performed in 65 (92%) trials, of which

72% (47/65) was reported as planned; 82% (53/65) reported stopping rules. Planned interim analyses were based on number of events (20/65, 31%), accrual targets (18/65, 28%) or time of study conduct (5/65, 8%). In 24 cases (37%), the trigger for the interim analysis was unspecified. Most trials describing stopping rules used O'Brien Fleming methodology (27/53, 51%) or specified a numeric threshold (20/53, 38%). Amongst the 30 trials which specified threshold *p*-value for study closure, the median *p*-value was 0.008 (range 0.00025–0.05).

Results of the RCTCEB are shown in Table 3. Trials on average accrued 75% of the planned sample size. Thirty-five percent of RCTs (21/60 evaluable RCTs) accrued less than 60% and 27% (16/60) accrued more subjects than originally planned; the latter trials were considered stopped early due to an interim analysis done after the completion of accrual but before mature follow-up data and/or targeted number of events were available. Comparisons of the observed versus planned number of events (as incorporated into a power calculation) were limited because these data were available for only six trials; amongst these, the mean observed number of events at the time of study closure was only 42% of expected (range 23–53%). In each of the 25 trials reporting a HR for the primary end-point, the *p*-value was <0.05; the mean HR was 0.50 (range 0.20–0.79).

Analyses of RCTCEB methodology are shown in Table 4. Trial methodology and reporting have improved between 1985–1999 and 2000–2008; a greater proportion of RCTCEB now reports a prospectively determined primary end-point (41/43 [95%] versus 14/28 [50%]; $p<0.001$), involvement of DSMB (36/43 [84%] versus 8/28 [29%]; $p = 0.001$), a planned interim analysis (38/43 [88%] versus 16/28 [57%]; $p = 0.004$) and explicit stopping rules (40/43 [93%] versus 14/28 [50%]; $p = 0.004$). Contemporary RCTCEB are also enrolling a greater proportion of patients relative to the targeted sample size (mean 81% versus 63%; $p = 0.006$). Cooperative group trials were less likely to report using an ITT analysis as compared with non-cooperative group sponsorship (18/41 [44%] versus 24/30 [80%]; $p = 0.003$); other differences in design or analysis were not detected. RCTCEB sponsored by industry were more likely than trials not sponsored by industry to explicitly state the primary end-point (22/22 [100%] versus 33/49 [67%]; $p = 0.002$), close with a greater proportion of patients accrued

Table 3 – Results of RCTs closed early for benefit (*n* = 71).

	Mean, median (range)
Patients reported compared to planned sample size	
Patients reported/planned sample size (*n* = 60)	0.75, 0.76 (0.26, 1.23)
Ratio	
0.2 – <0.4	4/60 (7%)
0.4 – <0.6	17/60 (28%)
0.6 – <0.8	14/60 (23%)
0.8 – <1.0	9/60 (15%)
⩾1.0	16/60 (27%)
Median follow-up (months)	
Number trials that report follow-up	
All trials (*n* = 44)	31.0, 30.5 (6.6, 82.8)
Early stage trials (*n* = 32)[a]	35.0, 34.3 (10.5, 82.8)
Advanced stage trials (*n* = 11)[a]	17.1, 17.6 (6.6, 32.4)
Number of events	
All trials (*n* = 25)	153, 144 (11, 475)
Observed events/events specified in power calculation (*n* = 6)	0.42, 0.42 (0.23, 0.53)
0.2 – <0.4	1 (17%)
0.4 – <0.6	5 (83%)
Timing of trial conduct (months)	41.2, 38.0 (7, 123)
Analysis to closure (*n* = 31)	2.2, 1.0 (0, 11)
DSMB meeting to closure (*n* = 12)	0.3, 0 (0, 2)
Closure to publication (*n* = 47)	32.4, 26.0 (1, 90)
Summary measures[b]	
Trials with HR	25/71 (35%)
Number trials with HR $p < 0.05$	25/25 (100%)
Distribution of HR	0.50, 0.51 (0.20, 0.79)
Trials with RR	3/71 (4%)

[a] For solid tumour trials early-stage refers to neoadjuvant/adjuvant therapy and advanced-stage refers to palliative therapy for metastatic disease. For haematologic malignancies, early-stage refers to first-line therapy and advanced-stage refers to trials in the relapsed or refractory settings.

[b] Reported for primary end-point. HR for RCTCEB favouring control arm has been converted to 1/HR to reflect consistent measurement of effect size across all RCTCEB.

Table 4 – Methodology of RCTCEB ($n = 71$).

	Year			Cooperative group			Industry sponsorship			Arm showing benefit			Follow-up reported		
	1985–1999 $n = 28$	2000–2008 $n = 43$	p^a	Yes $n = 41$	No $n = 30$	p^a	Yes $n = 22$	No $n = 49$	p^a	Experimental $n = 58$	Control $n = 13$	p^a	Yes $n = 23$	No $n = 48$	p^a
Primary EP	14 (50%)	41 (95%)	<0.001	29 (71%)	26 (87%)	0.239	22 (100%)	33 (67%)	0.002	47 (81%)	8 (62%)	0.271	18 (78%)	37 (77%)	1.000
Intention-to-treat (ITT)	12 (43%)	30 (70%)	0.029	18 (44%)	24 (80%)	0.003	18 (82%)	24 (49%)	0.010	34 (59%)	8 (62%)	1.000	11 (48%)	31 (65%)	0.205
Power calculated	20 (71%)	39 (91%)	0.051	32 (78%)	27 (90%)	0.218	21 (96%)	38 (78%)	0.089	48 (83%)	11 (85%)	1.000	20 (87%)	39 (81%)	0.739
DSMB	8 (29%)	36 (84%)	0.001	27 (66%)	17 (57%)	0.639	21 (96%)	23 (47%)	0.062	37 (64%)	7 (54%)	0.224	14 (61%)	30 (63%)	1.000
IA conducted	24 (86%)	41 (95%)	1.000	38 (93%)	27 (90%)	0.424	22 (100%)	43 (88%)	1.000	52 (90%)	13 (100%)	1.000	22 (96%)	43 (90%)	1.000
IA planned	16 (57%)	38 (88%)	0.004	32 (78%)	22 (73%)	0.639	20 (91%)	34 (69%)	0.156	47 (81%)	7 (54%)	0.030	18 (78%)	36 (75%)	1.000
Stopping rule	14 (50%)	40 (93%)	0.004	31 (76%)	23 (77%)	1.000	20 (91%)	34 (69%)	0.405	48 (83%)	6 (46%)	0.001	18 (78%)	36 (75%)	1.000
[b]Actual versus planned accrual	0.63 (0.36,1.05)	0.81 (0.26,1.23)	0.006	0.76 (0.37,1.23)	0.74 (0.26,1.16)	0.778	0.91 (0.26,1.16)	0.67 (0.34,1.23)	<0.001	0.76 (0.26,1.23)	0.71 (0.37,0.91)	0.411	0.80 (0.36,1.13)	0.72 (0.26,1.23)	0.303
$n = 60$															
[b]Months follow-up $n = 44$	31.88 (10.5,66)	30.35 (6.6,82.8)	0.780	32.63 (8.5,69)	28.35 (6.6,82.8)	0.436	22.45 (6.6,54.6)	34.56 (8.5, 82.8)	0.034	30.05 (6.6,69)	36.85 (10.5,82.8)	0.594	29.62 (8.5, 59.5)	31.75 (6.6,82.8)	0.701
[b]Number events $n = 25$	95.5 (13,368)	179.47 (11,475)	0.139	157.56 (16,475)	143.78 (11,415)	0.807	224.2 (18,475)	104.87 (11,415)	0.023	124.95 (11,475)	263.2 (203,415)	0.032	151 (18,475)	153 (11,415)	0.969

[a] Calculated using Fisher's exact test for proportions and *t*-test for means.
[b] Values shown reflect mean and range.

compared with the planned sample size (mean 91% versus 67%; $p < 0.001$) and observe a greater number of events at time (mean 224 versus 105; $p = 0.023$).

Our search for follow-up results identified articles for 18 RCTCEB and abstracts for an additional five RCTCEB. Comparative data for the trial primary end-point were available in 16/18 and 3/5 cases, repectively,. As shown in Table 5, the magnitude of effect and level of significance were very consistent from primary publication to follow-up report in all cases with a single exception. Amongst the 23 pairs evaluable for author conclusions, we only identified one study (4%) with a discordant conclusion.[11,12] RCTCEB that favoured the experimental arm were more likely than those favouring the control arm to have published follow-up reports (22/58 [38%] versus 1/13 [8%]; $p = 0.049$). We did not detect any differences between those trials associated with follow-up reports versus those that were not with respect to the reported parameters of methodological quality (Table 4).

4. Discussion

We reviewed 71 oncology RCTs published between 1985 and 2008 that were closed early for benefit. Several important findings have emerged. First, these reports are becoming increasingly common, are usually published in widely-read journals and are most likely to evaluate patients with breast, lung and haematologic cancers. Second, most trials closed before accrual has been completed. Third, we identified several aspects of study design and reporting that are deficient; 23% of reports did not explicitly identify the study primary end-point and only 59% included an ITT analysis of all randomised patients. Fourth, we found that only 62% described involvement of a DSMB. Fifth, up to 28% of RCTCEB was closed based on unplanned interim analyses.

Despite these limitations, we found reasons for reassurance. First, more recently published RCTCEB are associated with improved methodological rigour; planned interim analyses with formal boundaries for early trial closure are described in at least 88% of these trials published since 2000. Finally, while we could only identify follow-up reports for 32% of all RCTCEB (and 38% of those for which experimental therapy was deemed superior), study results and author conclusions were unchanged in all cases except for one RCT, suggesting that initial efficacy results are robust. The only example of a RCTCEB with inconsistent results on follow-up was RTOG-9413 in which the initial treatment effect favouring whole-pelvic over prostate-only radiotherapy disappeared with longer follow-up.[11,12]

Given the increasing frequency of RCTCEB, it is appropriate to consider risks and benefits associated with these reports. Provided that observed results are robust, early trial closure and reporting of results represent an appropriate action from scientific, ethical and resource utilisation perspectives. When these results come from a properly designed and analysed trial and confidently reject the trial's *a priori* null hypothesis (or confirm superiority of a control arm), continued trial conduct delays adoption of a superior treatment and unnecessarily continues to consume valu-

Table 5 – Comparison of primary end-point results for RCTCEB with subsequent reports[a].

References	Disease	N	Year of report	Results
Piccart-Gebhart et al.[21]	Breast	5081	2005	2 year DFS 77% versus 86%
			2007	3 year DFS 74% versus 81%
Goss et al.[31]	Breast	5187	2003	4 year DFS 87% versus 93% HR DFS 0.57, $p < 0.00001$
			2005	4 year DFS 90% versus 94% HR DFS 0.58, $p < 0.01$
			2008	4 year DFS 91% versus 94% HR DFS 0.68, $p = 0.0001$
Fenaux et al.[32]	Leukaemia	102	1993	1 year EFS 50% versus 79%
			2000	1 year EFS 53% versus 83%
Fisher et al.[33]	Breast	13,388	1998	RR incidence rate 0.51, $p < 0.00001$
			2005	RR incidence rate 0.57, $p < 0.001$
Herskovic et al.[34]	Oesophagus	129	1992	2 year OS 10% versus 38% MS 9 versus 13 mos
			1997	2 year OS 10% versus 36% MS 9 versus 14 mos
			1999	2 year OS 10% versus 36%
Lau et al.[35]	HCC	43	1999	3 year DFS 36% versus 75%
			2008	5 year DFS 32% versus 62%
Rosell et al.[36]	NSCLC	63	1994	MS 8 versus 26 mos, $p < 0.001$
			1999	MS 10 versus 22 mos, $p = 0.005$
Roth et al.[28]	NSCLC	60	1994	3 year OS 15% versus 56% MS 11 versus 64 mos, $p = 0.0013$
			1998	3 year OS 19% versus 43% MS 14 versus 21 mos, $p = 0.048$
FCCCLL[37]	Leukaemia	70	1986	2 year OS 44% versus 77%, $p = 0.0013$
			1989	3 year OS 28% versus 71%, $p = 0.0005$
Kaye et al.[38]	Ovary	165	1992	HR OS 0.53, $p = 0.003$
			1996	HR OS 0.68, $p = 0.043$
Richardson et al.[39]	Myeloma	669	2005	TTP 4 versus 6 mos
			2007	TTP 4 versus 6 mos
Motzer et al.[40]	RCC	750	2007	PFS 5 versus 11 mos
			2008	PFS 5 versus 11 mos
Geyer et al.[41]	Breast	324	2006	TTP 4 versus 8 mos HR TTP 0.49, $p < 0.001$
			2007	TTP 4 versus 6 mos HR TTP 0.57, $p = 0.00013$
Markman et al.[42]	Ovary	277	2003	PFS 21 versus 28 mos, $p = 0.035$
			2006	PFS 14 versus 22 mos, $p = 0.01$
Moertel et al.[43]	Colon	929	1990	HR OS 0.67, $p = 0.006$
			1995	HR OS 0.67, $p = 0.0007$
Tallman et al.[44]	Leukaemia	346	1997	3 year DFS 32% versus 67%, $p < 0.001$
			2002	5 year DFS 29% versus 69%, $p < 0.001$
Peters et al.[45]	Cervix	243	2000	4 year OS 71% versus 81%, $p = 0.007$
			2005	5 year OS 66% versus 80%, $p = N/A$
Morris et al.[46]	Cervix	388	1999	5 year OS 58% versus 73%, $p = 0.004$
			2004	5 year OS 52% versus 73%, $p < 0.0001$
Roach et al. [11,b]	Prostate	1292	2003	4 year PFS 46% versus 56%, $p = 0.014$
			2007	4 year PFS 54% versus 54%, $p = NS$

[a] DFS = disease-free survival; HR = hazard ratio; EFS = event-free survival; RR = risk ratio; OS = overall survival; MS = median survival; TTP = time to progression; PFS = progression-free survival; and N/A = not available.

[b] This trial had a factorial design; comparison of whole-pelvic versus prostate-only radiation was the one that crossed the interim monitoring boundary and is the one reported here. The results presented herein are taken from the analysis by Korn et al.[5] which were estimated from the interim monitoring boundary and from the survival curves presented in the follow-up report.[12]

able research resources. Potential risks of early closure have received considerable attention.[1–3,6] These mainly relate to statistical principles associated with performing early and multiple analyses. Comparative analyses based on a small number of events are intrinsically unstable due to random fluctuations and there is an increased probability of observing false results when multiple comparisons are performed.[13]

Through oversight of bodies such as a DSMB, the efficacy and adverse events of a trial must be monitored to protect the safety of enrolled subjects. Interim analyses are an essential component of these assessments. Interim analyses require greater levels of statistical significance than final analyses, and final analyses must account for multiple analyses.[14] Because preliminary data are potentially unstable, strategies to reduce risks of false-positive findings are needed as adoption of new treatments based on such results might subject future patients to therapy that is either ineffective or harmful. In our study, we were able to compare results of the primary end-point and author conclusions for 23 RCTCEB associated with published follow-up. While it is encouraging that discordant results were observed in only a single case, suggesting that benefits and risks of the superior therapy are preserved over time, we cannot exclude the possibility of bias in preferentially reporting trials that continue to demonstrate such benefits. Another potential risk occurs when early trial closure based on evaluation of one end-point compromises the ability to rigourously evaluate other end-points (such as quality of life, overall survival, and long-term toxicity), especially when cross-over of patients from one arm to another occurs. This risk is particularly important when the

trial's primary outcome is of a surrogate nature (e.g. progression or disease-free survival) and has an uncertain relationship with a more definitive health care policy-driving outcome (e.g. overall survival). Finally, RCT closure for benefit may result in altering the conduct of other similar studies (e.g. trial closure, compromise of accrual), thus precluding full testing of the hypotheses upon which these trials were based.

In order to maximise potential benefits and minimise potential risks associated with early RCT closure, fundamental biostatistical and methodological processes have been proposed.[6,7] However, we found that 23% of RCTCEB are reported without explicit definition of the primary end-point. This is consistent with the previous work by our group and others.[8,15] In order to minimise the chance of false positive results, interim analyses should be pre-planned, performed under the oversight of an independent DSMB, and have a more stringent threshold for statistical significance than the usual alpha of 0.05. While it is concerning that a substantial proportion of studies we identified were closed on the basis of unplanned analyses and appear to have been conducted without a DSMB, it is important to again note that improvements over time have been observed. Finally, robust follow-up data were not identified for 68% of trials in our cohort. Long-term follow-up of efficacy and toxicity results is essential so that clinicians can understand whether potential practice-changing results are durable.

The results of this study build upon work done by others. In their overview of 143 RCTs stopped early for benefit, Montori found that, on average, RCTs recruited 63% of the planned sample size and stopped after a median of 13 months follow-up, and when a median of 66 patients had experienced the end-point driving study termination.[1] The median risk ratio amongst these 143 RCTs was 0.53. Reporting of interim analysis statistical methodology was found to be poor. The authors concluded that RCTCEB often fail to report relevant information about the decision to stop early and observed implausibly large treatment effects and suggested that clinicians should view such results with skepticism. In a subgroup analysis of the 29 oncology RCTs within this cohort, Wilcox also described a significant correlation between fewer events and treatment effect ($r = 0.75$, $p < 0.0001$).[2] In a related study, Trotta analysed 25 oncology RCTCEB published between 1997 and 2007 and suggested that early termination may be driven by commercial interests and that only untruncated trials could provide high-quality evidence that should translate into clinical practice.[3]

As we have previously commented[16,17], the above analyses do not appear to account for the potential that appropriate methodological and statistical principles may have been applied and that early trial closure and reporting represent 'true-positive' rejection of a trial's null hypothesis. Korn's description of 27 NCI-funded co-operative group trials that were closed early for benefit supports the conclusion that early closure may be appropriate.[5] In that analysis, they found that amongst 18 trials for which follow-up was available, initial magnitudes of benefit were preserved in 17 (94%) trials; an accompanying editorial concluded that '...the system is working'.[18]

While our results are compatible with the work by Korn et al.[5], they are divergent from the recent study by Bassler et al.[4] who compared the treatment effect from 91 RCTs (including 10 oncology) closed early for benefit with that from meta-analyses of RCTs addressing the same question but not stopped early. They found that the pooled ratio of relative risks amongst the truncated RCTs versus matching non-truncated RCTs was 0.71 (96% CI 0.65–0.77). Furthermore, they found that the pooled effects in most (62%) non-truncated RCTs failed to demonstrate significant benefit. From an oncological perspective, this report is limited in power as it included only 10 RCTCEB. Furthermore, on review it appears that at least some of the non-truncated oncology RCTs used for comparative purposes included patients with different disease histology[19,20] and stages[21–24], and involved different chemotherapy drugs and regimens.[25–29] These differences raise important questions about whether their findings extend to RCTCEB in oncology.

Reasons for early trial closure may differ based on the stage of therapeutic development. Amongst the 83 RCTs closed early that were identified in our initial literature search (Fig. 1), 50 RCTs (50/83, 60%) were closed for benefit and 23 (23/83, 27%) were closed for futility. In their overview of randomised phase II trials, Lee and Feng[30] identified 45 trials that were closed early. Amongst these trials, only 13% were closed for benefit and 69% were closed for futility. This likely relates to the fact that agents undergoing phase III evaluation have already met a minimum efficacy bar in earlier trials and are, therefore, less likely to lead to futility than agents under study in the phase II setting.

Our study is the largest overview of oncology RCTs closed early for benefit. It is also the first to explore methodology of interim analyses and compare conclusions at time of trial closure with those observed with longer follow-up. Our study is limited by potentially having missed RCTCEB with our search strategy and basing our analyses of trial design and statistical methodology on reported information; it remains possible that some RCTCEB adhere to recommended design principles but do not report these details. Finally, our search for follow-up reports may be limited by publication bias and/or may underestimate the level of reporting since our search periods for RCTCEB and follow-up were both conducted as of 31st May 2008.

In summary, RCTCEB in oncology are increasingly common and quality of these trials has improved over time. Our results demonstrate that when reported, the conclusions of RCTCEB are preserved with longer follow-up. However, limitations in reporting of long-term follow-up remain and a potential for publication bias cannot be excluded. It is essential for trialists to properly design, analyse and report details of their studies so that other investigators, clinicians and policy-makers can estimate the validity of results when trials are closed early and carefully consider the implications of their adoption into practice.

Conflict of interest statement

None declared.

Acknowledgements

Dr. Booth is supported as a Cancer Care Ontario Chair in Health Services Research. The NCIC Clinical Trials Group,

including the salaries of Drs. Meyer and Tu, is supported by the Canadian Cancer Society Research Institute with funds provided by the Canadian Cancer Society. The authors acknowledge the input of Dr. Ben Djulbegovic in initial discussions related to this study protocol.

REFERENCES

1. Montori VM, Devereaux PJ, Adhikari NK, et al. Randomized trials stopped early for benefit: a systematic review. *JAMA* 2005;**294**(17):2203–9.
2. Wilcox RA, Djulbegovic B, Moffitt HL, Guyatt GH, Montori VM. Randomized trials in oncology stopped early for benefit. *J Clin Oncol* 2008;**26**(1):18–9.
3. Trotta F, Apolone G, Garattini S, Tafuri G. Stopping a trial early in oncology: for patients or for industry? *Ann Oncol* 2008;**19**(7):1347–53.
4. Bassler D, Briel M, Montori VM, et al. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. *JAMA* 2010;**303**(12):1180–7.
5. Korn EL, Freidlin B, Mooney M. Stopping or reporting early for positive results in randomized clinical trials: the National Cancer Institute Cooperative Group experience from 1990 to 2005. *J Clin Oncol* 2009;**27**(10):1712–21.
6. Girling D, Parmar M, Stenning S, Stephens R, Stewart L. *Clinical trials in cancer*. Oxford, UK: Oxford University Press; 2007.
7. Green SJ, Fleming TR, O'Fallon JR. Policies for study monitoring and interim reporting of results. *J Clin Oncol* 1987;**5**(9):1477–84.
8. Booth CM, Cescon DW, Wang L, Tannock IF, Krzyzanowska MK. Evolution of the randomized controlled trial in oncology over three decades. *J Clin Oncol* 2008;**26**(33):5458–64.
9. Booth CM, Le MA, Ding K, et al. Presentation of nonfinal results of randomized controlled trials at major oncology meetings. *J Clin Oncol* 2009;**27**(24):3938–44.
10. Ridker PM, Torres J. Reported outcomes in major cardiovascular clinical trials funded by for-profit and not-for-profit organizations: 2000–2005. *JAMA* 2006;**295**(19):2270–4.
11. Roach III M, DeSilvio M, Lawton C, et al. Phase III trial comparing whole-pelvic versus prostate-only radiotherapy and neoadjuvant versus adjuvant combined androgen suppression: Radiation Therapy Oncology Group 9413. *J Clin Oncol* 2003;**21**(10):1904–11.
12. Lawton CA, DeSilvio M, Roach III M, et al. An update of the phase III trial comparing whole pelvic to prostate only radiotherapy and neoadjuvant to adjuvant total androgen suppression: updated analysis of RTOG 94-13, with emphasis on unexpected hormone/radiation interactions. *Int J Radiat Oncol Biol Phys* 2007;**69**(3):646–55.
13. Fleming TR, Ellenberg S, DeMets DL. Monitoring clinical trials: issues and controversies regarding confidentiality. *Stat Med* 2002;**21**(19):2843–51.
14. Lan K, De Mets DL. Discrete sequential boundaries for clinical trials. *Biometrika* 1983;**70**:659–63.
15. Krzyzanowska MK, Pintilie M, Brezden-Masley C, Dent R, Tannock IF. Quality of abstracts describing randomized trials in the proceedings of American Society of Clinical Oncology meetings: guidelines for improved reporting. *J Clin Oncol* 2004;**22**(10):1993–9.
16. Pater JL, Goss P, Meyer R. Stopping trials for benefit can (sometimes) benefit patients. *J Clin Oncol* 2008;**26**(16):2787–8.
17. Bradbury P, Meyer R, Pater J, et al. Stopping a trial early in oncology: for patients or for industry? *Ann Oncol* 2009;**20**(2):395–6.
18. Sargent D. Early stopping for benefit in National Cancer Institute-sponsored randomized phase III trials: the system is working. *J Clin Oncol* 2009;**27**(10):1543–4.
19. Cooper JS, Guo MD, Herskovic A, et al. Chemoradiotherapy of locally advanced esophageal cancer: long-term follow-up of a prospective randomized trial (RTOG 85-01). Radiation Therapy Oncology Group. *JAMA* 1999;**281**(17):1623–7.
20. Slabber CF, Nel JS, Schoeman L, et al. A randomized study of radiotherapy alone versus radiotherapy plus 5-fluorouracil and platinum in patients with inoperable, locally advanced squamous cancer of the esophagus. *Am J Clin Oncol* 1998;**21**(5):462–5.
21. Piccart-Gebhart MJ, Procter M, Leyland-Jones B, et al. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N Engl J Med* 2005;**353**(16):1659–72.
22. Romond EH, Perez EA, Bryant J, et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *N Engl J Med* 2005;**353**(16):1673–84.
23. Slamon DJ, Leyland-Jones B, Shak S, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* 2001;**344**(11):783–92.
24. Marty M, Cognetti F, Maraninchi D, et al. Randomized phase II trial of the efficacy and safety of trastuzumab combined with docetaxel in patients with human epidermal growth factor receptor 2-positive metastatic breast cancer administered as first-line treatment: the M77001 study group. *J Clin Oncol* 2005;**23**(19):4265–74.
25. Freiha F, Reese J, Torti FM. A randomized trial of radical cystectomy versus radical cystectomy plus cisplatin, vinblastine and methotrexate chemotherapy for muscle invasive bladder cancer. *J Urol* 1996;**155**(2):495–9.
26. Lehmann J, Franzaring L, Thuroff J, et al. Complete long-term survival data from a trial of adjuvant chemotherapy vs control after radical cystectomy for locally advanced bladder cancer. *BJU Int* 2006;**97**(1):42–7.
27. Studer UE, Bacchi M, Biedermann C, et al. Adjuvant cisplatin chemotherapy following cystectomy for bladder cancer: results of a prospective randomized trial. *J Urol* 1994;**152**(1):81–4.
28. Roth JA, Fossella F, Komaki R, et al. A randomized trial comparing perioperative chemotherapy and surgery with surgery alone in resectable stage IIIA non-small-cell lung cancer. *J Natl Cancer Inst* 1994;**86**(9):673–80.
29. Depierre A, Milleron B, Moro-Sibilot D, et al. Preoperative chemotherapy followed by surgery compared with primary surgery in resectable stage I (except T1N0), II, and IIIa non-small-cell lung cancer. *J Clin Oncol* 2002;**20**(1):247–53.
30. Lee JJ, Feng L. Randomized phase II designs in cancer clinical trials: current status and future directions. *J Clin Oncol* 2005;**23**(19):4450–7.
31. Goss PE, Ingle JN, Martino S, et al. A randomized trial of letrozole in postmenopausal women after five years of tamoxifen therapy for early-stage breast cancer. *N Engl J Med* 2003;**349**(19):1793–802.
32. Fenaux P, Le Deley MC, Castaigne S, et al. Effect of all transretinoic acid in newly diagnosed acute promyelocytic leukemia. Results of a multicenter randomized trial. European APL 91 Group. *Blood* 1993;**82**(11):3241–9.
33. Fisher B, Costantino JP, Wickerham DL, et al. Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. *J Natl Cancer Inst* 1998;**90**(18):1371–88.

34. Herskovic A, Martz K, al-Sarraf M, et al. Combined chemotherapy and radiotherapy compared with radiotherapy alone in patients with cancer of the esophagus. *N Engl J Med* 1992;**326**(24):1593–8.
35. Lau WY, Leung TW, Ho SK, et al. Adjuvant intra-arterial iodine-131-labelled lipiodol for resectable hepatocellular carcinoma: a prospective randomised trial. *Lancet* 1999;**353**(9155):797–801.
36. Rosell R, Gomez-Codina J, Camps C, et al. A randomized trial comparing preoperative chemotherapy plus surgery with surgery alone in patients with non-small-cell lung cancer. *N Engl J Med* 1994;**330**(3):153–8.
37. French Cooperative Group on Chronic Lymphocytic Leukaemia. Effectiveness of "CHOP" regimen in advanced untreated chronic lymphocytic leukaemia. *Lancet* 1986;**1**(8494):1346–9.
38. Kaye SB, Lewis CR, Paul J, et al. Randomised study of two doses of cisplatin with cyclophosphamide in epithelial ovarian cancer. *Lancet* 1992;**340**(8815):329–33.
39. Richardson PG, Sonneveld P, Schuster MW, et al. Bortezomib or high-dose dexamethasone for relapsed multiple myeloma. *N Engl J Med* 2005;**352**(24):2487–98.
40. Motzer RJ, Hutson TE, Tomczak P, et al. Sunitinib versus interferon alfa in metastatic renal-cell carcinoma. *N Engl J Med* 2007;**356**(2):115–24.
41. Geyer CE, Forster J, Lindquist D, et al. Lapatinib plus capecitabine for HER2-positive advanced breast cancer. *N Engl J Med* 2006;**355**(26):2733–43.
42. Markman M, Liu PY, Wilczynski S, et al. Phase III randomized trial of 12 versus 3 months of maintenance paclitaxel in patients with advanced ovarian cancer after complete response to platinum and paclitaxel-based chemotherapy: a Southwest oncology group and gynecologic oncology group trial. *J Clin Oncol* 2003;**21**(13): 2460–645.
43. Moertel CG, Fleming TR, Macdonald JS, et al. Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. *N Engl J Med* 1990;**322**(6):352–8.
44. Tallman MS, Andersen JW, Schiffer CA, et al. All-trans-retinoic acid in acute promyelocytic leukemia. *N Engl J Med* 1997;**337**(15):1021–8.
45. Peters III WA, Liu PY, Barrett RJ, et al. Concurrent chemotherapy and pelvic radiation therapy compared with pelvic radiation therapy alone as adjuvant therapy after radical surgery in high-risk early-stage cancer of the cervix. *J Clin Oncol* 2000;**18**(8):1606–13.
46. Morris M, Eifel PJ, Lu J, et al. Pelvic radiation with concurrent chemotherapy compared with pelvic and para-aorticradiation for high-risk cervical cancer. *N Engl J Med* 1999;340(15):1137(Group experience from 1990 to 2005. *J Clin Oncol* 2009;**27**(10):1712-1721.