VOLUME 27 · NUMBER 10 · APRIL 1 2009

JOURNAL OF CLINICAL ONCOLOGY

STATISTICS IN ONCOLOGY

Stopping or Reporting Early for Positive Results in Randomized Clinical Trials: The National Cancer Institute Cooperative Group Experience From 1990 to 2005

Edward L. Korn, Boris Freidlin, and Margaret Mooney

From the Biometric Research Branch and the Clinical Investigations Branch, Cancer Therapy Evaluation Program, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, MD.

Submitted August 7, 2008; accepted November 7, 2008; published online ahead of print at www.jco.org on February 23, 2009.

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Corresponding author: Edward L. Korn, PhD, Biometric Research Branch, EPN-8129, National Cancer Institute, Bethesda, MD 20892; e-mail: korne@ctep.nci.nih.gov.

The Appendix is included in the full-text version of this article, available online at www.jco.org. It is not included in the PDF version (via Adobe® Reader®).

Published by the American Society of Clinical Oncology

0732-183X/09/2710-1712/$20.00

DOI: 10.1200/JCO.2008.19.5339

ABSTRACT

Randomized clinical trials are designed with stopping boundaries to guide data monitoring committees with their decision making concerning ongoing trials. In particular, when extremely positive results are seen and a boundary is crossed, the data monitoring committee may recommend releasing the results earlier to the public than at the definitive final analysis time specified in the protocol. For trials that are still accruing, this also means stopping accrual. Because the information about treatment efficacy is more limited in an early analysis than in a final analysis, questions have been raised about the appropriateness of incorporating early stopping for positive results in trial designs. In particular, there are concerns that treatment effects seen early may not be real or may be overly optimistic. To examine this issue, we collected information about treatment efficacy on National Cancer Institute Cooperative Group trials that were stopped early for positive results (information both at the time the trial was stopped/released and at times of further follow-up). Twenty-seven such trials were located. For 17 of 18 of these trials with sufficient follow-up information, the treatment effect was similar or only slightly smaller at last follow-up compared with the stopping/release time. We critically evaluate reasons why one might be concerned about early stopping for positive results. We conclude that for trials with well-designed interim monitoring plans, the ability to stop early for positive results is an important component of the trial design, allowing the public to benefit as soon as possible from the study conclusions.

J Clin Oncol 27:1712-1721. Published by the American Society of Clinical Oncology

INTRODUCTION

Interim monitoring plans, including formal guidelines for stopping trials early (stopping accrual and/or releasing results early) for compelling results, are a standard part of the designs of randomized clinical trials (RCTs). The monitoring guidelines are designed to limit the probability of a false-positive result (type I error) while allowing trials to stop early. Although the benefits to the public in releasing compelling positive trial results early are obvious, there have been concerns expressed[1-5] about the correctness of early stopping of RCTs for positive results. To address the concerns raised, we performed a review of all treatment RCTs performed by the National Cancer Institute (NCI) Cooperative Groups and published from 1990 to 2005, thus providing empiric data on the issues as has been suggested.[6] We also examined frequently cited reasons why one should be cautious about using early stopping for positive results and found some of these reasons to be correct but others lacking in statistical validity.

NCI COOPERATIVE GROUP TRIALS THAT STOPPED OR RELEASED RESULTS EARLY FOR POSITIVE RESULTS

We located NCI Cooperative Group phase III treatment RCTs whose accrual was stopped early or whose results were released early for positive results with the first publication appearing in the years 1990 to 2005; we included Canadian National Cancer Institute of Canada (NCIC) Clinical Trials Group trials partially supported by NCI. Using a list from PDQ (http://www.cancer.gov/cancertopics/pdq/cancerdatabase), all trial publications appearing in *Journal of Clinical Oncology*, *Journal of the National Cancer Institute*, *New England Journal of Medicine*, *The Lancet*, or *Blood* or as an American Society of Clinical Oncology or American Society of Hematology abstract were examined for evidence of early stopping/release for positive results. We located 27 such trials (Table 1);[7-60] it is possible that a small number of trials were missed if they were not published in the searched publications or if no mention was made of the early stopping/

Table 1. Descriptions of NCI Cooperative Group Phase III Treatment Trials That Stopped or Released Results Early for Positive Results, With Primary Publication Appearing Between 1990 and 2005

Trial Identifier (PDQ)	Trial Description and Reference Numbers*	Trial End Point†	Year Trial Was Stopped/Reported	% of Patients Accrued When Trial Stopped/Reported
NCCTG-844652	Phase III pilot of adjuvant therapy with levamisole *v* levamisole plus 5-fluorouracil for resectable adenocarcinoma of the colon[7,8]	OS	1989	100
RTOG-8501	Phase III comparison of radiotherapy alone *v* radiotherapy plus combination chemotherapy with CACP/FU in patients with localized carcinoma of the esophagus[9,10]	OS	1990	86
POG-9006	Phase III comparison of intensification with MP/MTX *v* alternating MP/MTX, VM-26/ARA-C, and DNR/ARA-C/VCR/PRED/PEG-ASP following induction with PRED/VCR/ASP/DNR in children with higher risk ALL[11,12]	CCR	1994	97
EST-3189	Phase III randomized comparison of CAF (CTX/ADR/FU) *v* a 16-week multidrug regimen (CTX/ADR/VCR/MTX/FU) as adjuvant therapy in node-positive patients with receptor-negative breast cancer[13,14]	OS	1994	100
CLB-9011	Phase III comparison of CLB *v* FAMP in previously untreated patients with intermediate- and high-risk (Rai stage I-IV) B-cell CLL[15,16]	CR	1995	100‡
SWOG-8892	Phase III study of radiotherapy with *v* without concurrent CDDP followed by CDDP/FU for previously untreated stage III/IV carcinoma of the nasopharynx[17-19]	PFS	1995	71
E-2491	Phase III randomized trial of tretinoin *v* ARA-C/DNR as induction therapy and of tretinoin *v* observation as maintenance therapy for patients with previously untreated promyelocytic leukemia[20]	DFS	1995	94
CCG-1882	Phase III treatment of poor-prognosis childhood ALL (excluding infants and patients with lymphoma-leukemia and FAB L3 blasts), including a randomized comparison of standard *v* augmented BFM in late responders[21]	EFS	1996	100
SWOG-8814	Phase III randomized comparison of adjuvant therapy with TMX *v* CAF (CTX/DOX/FU) plus concurrent or delayed TMX in postmenopausal women with node- and receptor-positive breast cancer[22,23]	DFS	1997	100
SWOG-8797	Phase III randomized comparison of radiotherapy with *v* without continuous-infusion FU/bolus CDDP after radical hysterectomy and node dissection in high-risk patients with stages IA2, IB, and IIA carcinoma of the cervix[24,25]	OS	1998	100
CLB-9344	Phase III randomized study of adjuvant cyclophosphamide/doxorubicin comparing standard- *v* intermediate- *v* high-dose doxorubicin, with *v* without subsequent paclitaxel, in women with node-positive breast cancer[26-28]	DFS	1998	100
RTOG-9001	Phase III comparison of pelvic irradiation with concurrent CDDP/5-FU *v* pelvic and para-aortic irradiation without chemotherapy in patients with high-risk carcinoma of the cervix[29,30]	OS	1998	100
CCG-5942	Phase III study of adjuvant low-dose involved-field radiotherapy *v* no adjuvant therapy in children with Hodgkin's disease in CR after chemotherapy assigned by clinical stage[31]	EFS	1998§	77
SWOG-9133	Phase III randomized trial of subtotal irradiation with *v* without DOX/VBL in patients with stage IA/IIA Hodgkin's disease[32,33]	FFS	2000	83
RTOG-9413	Phase III randomized study of whole pelvic irradiation followed by a cone-down boost to the prostate *v* prostate irradiation only and of neoadjuvant *v* adjuvant FLUT/ZDX for adenocarcinoma of the prostate[34-36]	PFS	2001	100
SWOG-S9701	Phase III randomized trial of 12 months *v* 3 months of paclitaxel in patients with advanced ovarian, fallopian tube, or primary peritoneal cancer in complete remission after platinum/paclitaxel–based chemotherapy[37,38]	PFS	2001	66
NCCTG-N9741	Phase III randomized study of combinations of oxaliplatin, fluorouracil, leucovorin calcium, and irinotecan as initial therapy in patients with advanced adenocarcinoma of the colon and rectum[39,40]	PFS	2002	100
NCIC-MA17	Phase III randomized study of letrozole *v* placebo in postmenopausal women with primary breast cancer who have completed at least 5 years of adjuvant aromatase inhibitor[41,42]	DFS	2003	100
E-1496	Phase III randomized study of standard therapy followed by maintenance therapy with rituximab (IDEC-C2B8 monoclonal antibody) or observation in patients with stage III or IV low-grade non-Hodgkin's lymphoma[43,44]	PFS	2003	100
E-E1A00	Phase III randomized study of dexamethasone with or without thalidomide in patients with newly diagnosed multiple myeloma[45,46]	RR	2003	100
CCG-1961	Phase III randomized study of treatment based on response to induction chemotherapy in patients with higher risk childhood acute lymphocytic leukemia: standard *v* augmented BFM regimen with standard *v* prolonged intensification for rapid early responders and doxorubicin *v* idarubicin and cyclophosphamide with delayed intensification for slow early responders[47,48]	EFS	2003	100
E-3200	Phase III randomized study of oxaliplatin, fluorouracil, and leucovorin calcium with or without bevacizumab *v* bevacizumab only in patients with previously treated advanced or metastatic colorectal adenocarcinoma[49-51]	OS	2004	100
ECOG-2997	Phase III randomized study of fludarabine with or without cyclophosphamide in patients with previously untreated chronic lymphocytic leukemia[52,53]	CR	2004	100

(continued on following page)

Table 1. Descriptions of NCI Cooperative Group Phase III Treatment Trials That Stopped or Released Results Early for Positive Results, With Primary Publication Appearing Between 1990 and 2005 (continued)

Trial Identifier (PDQ)	Trial Description and Reference Numbers*	Trial End Point†	Year Trial Was Stopped/Reported	% of Patients Accrued When Trial Stopped/Reported
NSABP-B-31/ NCCTG-N9831	Phase III randomized study of doxorubicin and cyclophosphamide followed by paclitaxel with or without trastuzumab in women with node-positive breast cancer that overexpresses HER-2/Phase III randomized study of doxorubicin plus cyclophosphamide followed by paclitaxel with or without trastuzumab in women with HER-2–overexpressing node-positive or high-risk node-negative breast cancer[54,55]	DFS	2005	81
ECOG-4599	Phase II/III randomized study of paclitaxel and carboplatin with or without bevacizumab in patients with advanced, metastatic, or recurrent non–squamous cell non–small-cell lung cancer[56]	OS	2005	100
ECOG-2100	Phase III randomized study of paclitaxel with or without bevacizumab in patients with locally recurrent or metastatic breast cancer[57,58]	PFS	2005	100
NCIC-MA21	Phase III randomized study of adjuvant cyclophosphamide, epirubicin, and fluorouracil *v* cyclophosphamide, epirubicin, filgrastim (G-CSF), and epoetin alfa followed by paclitaxel *v* cyclophosphamide and doxorubicin followed by paclitaxel in premenopausal or early postmenopausal women with previously resected node-positive or high-risk node-negative stage I-IIIB breast cancer[59,60]	DFS	2005	100

Abbreviations: NCI, National Cancer Institute; NCCTG, North Central Cancer Treatment Group; OS, overall survival; RTOG, Radiation Therapy Oncology Group; CACP, cisplatin; FU, fluorouracil; POG, Pediatric Oncology Group; MP, mercaptopurine; MTX, methotrexate; VM-26, teniposide; ARA-C, cytarabine; DNR, daunorubicin; VCR, vincristine; PRED, prednisone; PEG, pegylated; ASP, asparaginase; ALL, acute lymphoblastic leukemia; CCR, continuous complete remission; CTX, cyclophosphamide; ADR, doxorubicin; CLB, chlorambucil; FAMP, fludarabine; CLL, chronic lymphocytic leukemia; CR, complete response; SWOG, Southwest Oncology Group; CDDP, cisplatin; PFS, progression-free survival; DFS, disease-free survival; CCG, Children's Cancer Group; FAB, French-American-British; BFM, Berlin-Frankfurt-Muenster; EFS, event-free survival; TMX, tamoxifen; DOX, doxorubicin; VBL, vinblastine; FFS, failure-free survival; FLUT, flutamide; ZDX, goserelin; NCIC, National Cancer Institute of Canada; RR, response rate; ECOG, Eastern Cooperative Oncology Group; NSABP, National Surgical Adjuvant Breast and Bowel Project; HER-2, human epidermal growth factor receptor 2; G-CSF, granulocyte colony-stimulating factor.

*Reference numbers are given for references that were used to complete the trial information for Tables 1 to 3. Additional information on the trials was obtained from personal communications, as follows: EST-3189, E-1496, E-3200, ECOG-2997 (R.J. Gray, personal communication, July 2008); CLB-9011 (B. Peterson, personal communication, August 2008); SWOG-8892 (M. LeBlanc, personal communication, July 2008); SWOG-8814 (W. Barlow, personal communication, July 2008); SWOG-8797 (P.Y. Liu, personal communication, June 2008); RTOG-9001 (K. Winter, personal communication, July 2008); CCG-5942 (R. Sposto, personal communication, July 2008, of interim report on study 5942 prepared for the CCG Data and Safety Monitoring Committee, October 12, 1998); CCG-1961 (N.L. Seibel, personal communication, May 2008); and CCG-1882 (M. Devidas, personal communication, August 2008).

†Complete definitions of the trial end points are given in Appendix Table A1.

‡This is the percentage of the original target sample size; the primary end point was changed and the sample size was increased after the interim analysis when the results crossed the boundary in 1993. At that time, the percentage of accrual was 59%.

§Random assignment stopped in 1998; as per protocol, results were reported later (2002).

release in the publications. (Once a relevant trial was located, all publications associated with that trial were examined.) The end points that crossed interim monitoring boundaries for the 27 trials were overall survival (OS; seven trials), progression-free survival (PFS; six trials), disease-free survival (DFS; six trials), event-free survival (EFS; three trials), complete response rate (two trials), failure-free survival (one trial), continuous complete remission rates (one trial), and response rate (one trial). End point definitions can be found in Appendix Table A1 (online only). Accrual was complete at the time of stopping/release for 70% of the trails (19 of 27 trials) and ranged from 66% to 97% complete for the other eight trials.

Table 2 lists efficacy statistics on these trials at their times when they crossed their interim monitoring boundary, they were first published, and further follow-up information was published (when available). Efficacy data are given as reported in the publications (eg, hazard ratio or the 3-year survival in each arm), but all *P* values given here are two-sided. The follow-up information is an attempt to assess, in hindsight, how accurate or inaccurate the early positive results were. Regardless of these findings, there is no intended implication that the data monitoring committees making these particular stopping decisions made incorrect decisions given the protocol interim monitoring guidelines and the information they had at the time. We focus here on the results at the time of interim monitoring boundary crossing and at the last follow-up available (which could be when the results were first published if there was no additional follow-up). When the trial results crossed the boundary, the ratio of observed events to events required at the final analysis (information fraction) ranged from 15% to 90%, with a median of approximately 60%.

Although all 27 trials met their prespecified study objectives when their positive results crossed their boundaries, to provide a summary characterization of Table 2, we first focused on the 18 trials that had follow-up information of at least 80% (an arbitrary figure representing trials for which planned final analysis results could be considered available). For 14 of the 18 trials (North Central Cancer Treatment Group NCCTG-844652; Radiation Therapy Oncology Group RTOG-8501; Cancer and Leukemia Group B CLB-9011; Eastern Cooperative Oncology Group ECOG-2491; Children's Cancer Group CCG-1882; RTOG-9001; NCCTG-9741; Eastern Cooperative Oncology Group E1496; Eastern Cooperative Oncology Group E-E1A00; CCG-1961; E3200; E2997; National Surgical Adjuvant Breast and Bowel Project B NSABP-B-31/NCCTG-N9831; ECOG-4599), the treatment effect was similar at early stopping/release and last follow-up. For Southwest Oncology Group SWOG-8814 and Eastern Cooperative Oncology Group ECOG-2100, the treatment effect became slightly smaller, although with the same statistical significance. For Eastern Cooperative Oncology Group EST-3189, the treatment effect became slightly smaller with the statistical significance much weaker (although still statistically significant based on the protocol specification). For RTOG-9413, the treatment effect disappeared; at the time of early release, 90% of the required events

Table 2. Statistics on NCI CTEP-Sponsored Trials Whose Accrual Was Stopped or Whose Results Were Released Early for Positive Results From 1990 to 2005

	Results When Trial Crossed Boundary			Results When Trial First Published			Follow-Up Results of Trial		
Trial Identifier	% Information	Treatment Effect	*P*	% Information	Treatment Effect	*P*	% Information	Treatment Effect	*P*
NCCTG-844652*	60	HR = 0.67	.0064		Same as when trial stopped		89	HR = 0.67	.0007
RTOG-8501†	~60	HR = ~0.49	.0045	87	2-year OS: 38% *v* 10%	< .001	108	2-year OS: 36% *v* 10%	< .001
POG-9006‡	~38	2-year CCR: 82% *v* 70.8%	.0016	~54	2-year CCR: 84% *v* 75%	.006	77	4-year CCR: 70.6% *v* 64%	.22
EST-3189§	62	3-year OS: 84% *v* 73%	.0050		Same as when trial stopped		90	4-year OS: 78.1% *v* 71.4%	.10
CLB-9011‖	34	CR rates: 30% *v* 2%	.00014	78	CR rates: 33% *v* 8%	$< 10^{-5}$	117	CR rates: 20% *v* 4%	$< 10^{-5}$
SWOG-8892	26	HR = 0.26	< .0001		Same as when trial stopped		50	HR = 0.31	< .001
ECOG-2491¶	50	1-year DFS: 92% *v* 57%	< .0001		Same as when trial stopped		> 100	5-year DFS: 69% *v* 29%	< .0001
CCG-1882	68	4-yr EFS: 75.4% *v* 57.2%	.0013	80	5-year EFS: 75.0% *v* 55.0%#	< .001	—	—	—
SWOG-8814	81	HR = 0.66	.002		Same as when trial stopped		120	HR = 0.76	.002
SWOG-8797	~60	HR = 0.45**	0.006**	~60-63††	HR = 0.50	.01	63	HR = .51‡‡	.007‡‡
CLB-9344§§	25	HR = 0.79	.013		Same as when trial stopped		59	HR = 0.83	.0023
RTOG-9001	58	5-year OS: 73% *v* 58%	.0027	59	5-year OS: 73% *v* 58%	.004	81	5-year OS: 73% *v* 52%	< .0001
CCG-5942	34‖‖	HR = 0.27	.0048	64	HR = 0.59	.057¶¶			
SWOG-9133	46	3-year FFS: 93% *v* 81%	< .001		Same as when trial stopped		48	3-year FFS: 94% *v* 81%	< .001
RTOG-9413##	90	4-year PFS: 56% *v* 46%	.014		Same as when trial stopped		156	4-year PFS: 54% *v* 54%	NS
SWOG-S9701	15	HR = 0.43	.0046		Same as when trial stopped		62	HR = 0.70	.008
NCCTG-N9741***	81	HR = 0.71	.0009		Same as when trial stopped		98	HR = 0.74	.0014
NCIC-MA17	40	HR = 0.57	.00008		Same as when trial stopped		48	HR = 0.58	< .001
E-1496	49	HR = 0.42	.00016	59	HR = 0.5	.00006	83†††	HR = 0.38	$< 10^{-5}$
E-E1A00	59	RR: 80% *v* 53%	.0046		Same as when trial stopped		108	RR: 63% *v* 41%	.0034
CCG-1961‡‡‡	84	5-year EFS: 78.8% *v* 69.8%	.0198	91	5-year EFS: 80% *v* 71%	.01	124	5-year EFS: 81.2% *v* 71.7%	< .001
E-3200§§§	90	HR = 0.74	.0024		Same as when trial stopped		111	HR = 0.75	.0011
ECOG-2997	79	CR rates: 22.9% *v* 5.8%	.0008	98	CR rates: 22.4% *v* 5.8%	.0002	107	CR rates: 23.4% *v* 4.6%	$< 10^{-5}$
NSABP-B-31/ NCCTG-N9831‖‖‖	55	HR = 0.48	< .0001		Same as when trial stopped		87	HR = 0.49	< .0001
ECOG-4599	72	HR = 0.78	.0076	100	HR = 0.79	.003	—	—	—
ECOG-2100	65	HR = 0.50	< .001	114	HR = 0.60	< .001	—	—	—
NCIC-MA21¶¶¶	58	HR = 0.60	.0006		Same as when trial stopped		—	—	—

Abbreviations: NCI, National Cancer Institute; CTEP, Cancer Therapy Evaluation Program; NCCTG, North Central Cancer Treatment Group; HR, hazard ratio; RTOG, Radiation Therapy Oncology Group; OS, overall survival; POG, Pediatric Oncology Group; CCR, continuous complete remission; CR, complete response; SWOG, Southwest Oncology Group; ECOG, Eastern Cooperative Oncology Group; DFS, disease-free survival; CCG, Children's Cancer Group; EFS, event-free survival; FFS, failure-free survival; PFS, progression-free survival; NS, not significant; NCIC, National Cancer Institute of Canada; RR, response rate; NSABP, National Surgical Adjuvant Breast and Bowel Project.

*NCCTG-844652: We consider only the analysis of the stage C patients here (which was prespecified). The trial was stopped because of differences in the levamisole plus fluorouracil arm compared with the observation arm, which are the results reported here.

†RTOG-8501: Information percentage for when the trial crossed the boundary was estimated based on length of accrual/follow-up and observed survival rates. Treatment effect HR for when the trial crossed the boundary was estimated based on *P* value and estimated number of events.

‡POG-9006: Information percentage for when trial crossed boundary and when trial was first published was estimated based on accrual/follow-up and observed CCR. In the follow-up period, treatment effect 2-year CCR rates were 85.2% and 80.4%.

§ EST-3189: OS was one of the two coprimary end points and the one that crossed the interim monitoring boundary. In the follow-up period, the treatment effect–estimated 3-year OS rates from Figure 2 are 83% *v* 77%. Regarding the *P* value, the protocol for this trial specified a one-sided 5% level test, which corresponds to a two-sided 10% level test.

‖CLB-9011: This was initially a three-armed trial, with the experimental fludarabine plus chlorambucil arm later dropped. The results presented here are for the comparison of fludarabine *v* chlorambucil (control). The *P* value used for the interim analysis for when trial crossed boundary was $P = .0005$ based on a comparison of both experimental arms combined compared with the control arm.

¶ECOG-2491: The results reported here are for the induction therapy comparison (there was also a maintenance treatment randomization in this trial). For the follow-up period, the estimated 1-year DFS rates from Figure 1 of Tallman et al[20] are 88% *v* 57%.

#CCG-1882: Regarding treatment effect results when the trial was first published, the estimated 4-year EFS rates from Figure 1 of Nachman et al [21] are 77% *v* 60%.

**SWOG-8797: HR and *P* value were unadjusted for stratification variables. The values adjusted for the random assignment stratification variables are 0.49 and $P = .02$.

††SWOG-8797: The information fraction is not given in the first publication but must be between when the trial was stopped and the later follow-up.

(continued on following page)

Table 2. Statistics on NCI CTEP-Sponsored Trials Whose Accrual Was Stopped or Whose Results Were Released Early for Positive Results From 1990 to 2005 (continued)

‡‡SWOG-8797: HR and *P* value were adjusted at follow-up for random assignment stratification factors.
§§CLB-9344: This trial had a factorial design; comparison of ± paclitaxel was the one that led to the release of the data and is the one reported here. HR for when trial crossed boundary was estimated based on *P* value and percent information.
||||CCG-5942: Per protocol, the random assignment was stopped when the boundary was crossed, but the results were not released until there was further follow-up.
¶¶CCG-5942: This result would be considered statistically significant because the trial was designed with a one-sided type I error of 0.10, which corresponds to a two-sided type I error of 0.20.
##RTOG-9413: This trial had a factorial design; comparison of whole-port *v* prostate-only radiation was the one that crossed the interim monitoring boundary and is the one reported here. The percent information for when trial crossed boundary was estimated from interim monitoring boundary. The 4-year PFS results in the follow-up period were estimated from curves in Figure 2A of Lawton et al.[35]
***NCCTG-9741: This trial had three arms; fluorouracil/leucovorin + oxaliplatin *v* fluorouracil/leucovorin + irinotecan is the comparison that crossed the interim monitoring boundary and is the one considered here.
†††E1496: Information is estimated based on CI width for the HR.
‡‡‡CCG-1961: We are considering the analysis of the intensity of treatment question in the rapid early responder subgroup, which is the one that crossed an interim analysis boundary.
§§§E-3200: Arms being discussed here are the infusional fluorouracil, leucovorin, and oxaliplatin (FOLFOX4) arm (control arm) and FOLFOX4 + bevacizumab arm; the bevacizumab-alone arm closed early for negative results.
||||||NSABP-B-31/NCCTG-N9831: The results reported here are from the combined analysis of the trials NSABP-B-31 and NCCTG-N983, which are considered one trial for the purposes of discussion here.
¶¶¶NCIC-MA21: This was a three-arm trial; results are given for epirubicin-cyclophosphamide/paclitaxel *v* doxorubicin-cyclophosphamide/paclitaxel as experimental and control treatments, respectively.

were observed, and the 4-year PFS rates were 56% *v* 46% ($P = .014$). When long-term follow-up became available (156% of the required events), the 4-year PFS rates were 54% in both arms. Because in this trial the results were released so close to the protocol-specified final analysis, one can argue that this is an illustration of a study with the final analysis reversed by longer follow-up.

Although not one of the 18 trials with at least 80% information follow-up, the release of data in POG-9006 (with accrual 97% complete) may seem problematic. The trial reported 2-year EFS rates of 84% *v* 75% ($P = .006$) with 54% information (when the data crossed the interim monitoring boundary, the 2-year EFS rates were 82% *v* 70.8%, $P = .0016$, with 38% information). With further follow-up and 77% information, the reported 4-year EFS rates were 70.6% *v* 64% ($P = .21$). We will return to discussion of this trial in the next section. For the other eight trials that did not have at least 80% information follow-up, one trial had no further follow-up (NCIC-MA21), and the other seven trials retained statistical significance, with two having a smaller treatment effect (CCG-5942 and SWOG-S9701) and five showing a similar treatment effect (SWOG-8892, SWOG-8797, CLB-9344, SWOG-9133, and NCIC-MA17).

STATISTICAL AND DESIGN ISSUES INVOLVING EARLY STOPPING/RELEASE

We discuss a number of reasons that have been used to suggest that early stopping for extremely positive results may not be appropriate.

Choice of Primary End Point

The designated primary end point of an RCT is the one that is used to make the definitive statement concerning treatment effectiveness. There can be controversy about the appropriate primary end point, with different end points yielding different required sample sizes for the trial. For example, a trial that demonstrates a positive treatment effect at its conclusion for PFS may not have a sufficient number of deaths at that time to evaluate conclusively OS benefits. The possibility of early stopping exacerbates this potential problem in that there may be little information available about the nonprimary end points if the trial is stopped early. For example, Cannistra[61] questioned the decision to close and report early the results of SWOG-S9701 and NCIC-MA17 based on PFS and DFS end points, respectively, when the OS data were immature. If one believes that an improvement in a non-OS end point results in direct patient benefit, then there should be little argument against stopping a trial early based on extremely positive results for that end point. For example, the NCIC-MA17 investigators[62] considered DFS an important clinical end point for their adjuvant breast cancer setting. However, sometimes a non-OS end point is used not because it directly represents patient benefit, but because it is a surrogate for OS. As a surrogate for OS, it may have more statistical power because events accumulate faster and because it may be less susceptible to potential confounding by treatment crossovers to the experimental arm after non-OS events in the control arm. In this case, it is not as clear that one would need to stop a trial early for positive non-OS treatment effects, unless the surrogate is uniformly accepted in the clinical community.

When a non-OS primary end point does not directly represent clinical benefit, a reasonable strategy is to use OS for the interim analysis for positive effects even though the primary end point is different. To do this, one would have to be comfortable continuing a trial that showed extremely positive results in the non-OS primary end point provided that OS differences were not large. Other possibilities include requiring extremely positive results for early stopping/release or starting the interim monitoring for positive effects at a late enough time point that accrual will be complete or almost complete. This may allow evaluation of the OS effect with further follow-up. For example, in Table 3, we see that seven of 16 trials that stopped based on non-OS end points in Table 2 eventually achieved OS treatment effects that were large enough and precise enough to attain statistical significance ($P < .05$); four trials did not have OS results available. A potentially useful strategy to ameliorate the dilution of treatment effect as a result of crossovers is to censor the OS data of the control arm patients at the time when positive results are released; see Bukowski et al[63] for an example.

Table 3. Latest OS Results for Trials in Table 2 Whose Primary End Point Was Not OS

Trial Identifier	Treatment Effect	95% CI for HR or Difference in Rates	*P*
CLB-9011	Median OS: 66 *v* 56 months	Not applicable	.10
SWOG-8892	5-year OS: 67% *v* 37%	Not available	.001
E2491	5-year OS: 69% *v* 45%; difference = 24%	13.4% to 34.6%	.0001
SWOG-8814	HR = 0.83	0.69 to 0.99	.04
CLB-9344	HR = 0.82	0.71 to 0.95	.0064
CCG-5942	3-year OS: 98% *v* 99%; difference = −1%	−3.3% to 1.3%	.90
RTOG-9413	4-year OS: 84.7% *v* 84.3%; difference = 0.4%	−4.6% to 5.4%	.94
SWOG-S9701	HR = 0.84	0.61 to 1.16	.30
NCCTG-9741	HR = 0.66	0.54 to 0.82	.0001
NCIC-MA17	HR = 0.82	0.57 to 1.19	.3
ECOG-1496	HR = 0.51	0.25 to 1.04	.06
CCG-1961	HR = 0.64	0.47 to 0.87	.005
E2997	2-year OS: 79% *v* 80%; difference = −1%	−14.3 to 12.3*	.69
NSABP-B-31/NCCTG-N9831	HR = 0.63	0.49 to 0.81	.0004
ECOG-2100	HR = 0.88	0.74 to 1.05	.16
NCIC-MA21	47 *v* 65 deaths	Not applicable	.09†

NOTE. No OS results are available for POG-9006, CCG-1882, E-E1A00, or SWOG-9133.
Abbreviations: OS, overall survival; HR, hazard ratio; SWOG, Southwest Oncology Group; CCG, Children's Cancer Group; RTOG, Radiation Therapy Oncology Group; NCCTG, North Central Cancer Treatment Group; NCIC, National Cancer Institute of Canada; ECOG, Eastern Cooperative Oncology Group; NSABP, National Surgical Adjuvant Breast and Bowel Project.
*Based on a binomial approximation using a Peto effective sample size.
†Derived using a Poisson approximation for the numbers of deaths.

Crossing Hazards

Depending on the shape of the experimental and control treatment survival curves, early release of data can lead to different conclusions than with additional follow-up. Figure 1 displays hypothetical curves, with the experimental treatment being better on average. Note that although the control treatment curve drops faster than the experimental treatment curve during the first 5 years, the opposite is true for years 6 to 10. This is known as a case of crossing hazards and does not imply that the experimental treatment is worse than the control treatment in the later years (see the Appendix, online only). An important implication of crossing hazards for a conventionally designed RCT is that the trial may have more power to reject the null hypothesis with less follow-up. For example, if the true survival curves are as displayed in Figure 1, then an RCT randomly assigning 750 patients per arm accruing uniformly over 4 years with 2 years of follow-up would have 85% power to reject the null hypothesis (one-sided type I error = 0.025, log-rank statistic). (All power calculations are derived from simulation of 10,000 data sets.) The same trial with 5 years of follow-up would have 54% power. (In special circumstances where one expects the survival curves to come together, alternatives to the log-rank test that weight the earlier data more heavily may be appropriate.)

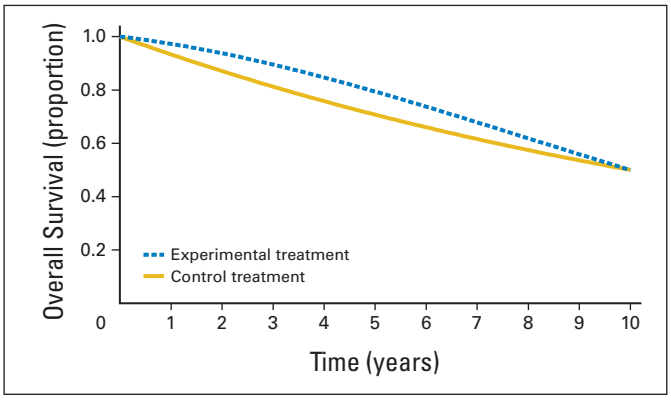

Fig 1. Theoretical survival curve data with crossing hazards at 5 years.

An implication of crossing hazards to the topic of early stopping is that an early highly statistically significant result leading to stopping the trial may become less statistically significant (or even not statistically significant) with further follow-up. (This could also happen with additional long-term follow-up of a trial that was not stopped early.) For example, suppose the RCT described earlier with 5 years of follow-up had an interim analysis after 2 years of follow-up that would release the trial results early if $P < .0025$. This would happen 57% of the time if the true survival curves were as in Figure 1, and with 3 years of additional follow-up, 22% of these times the results would no longer be statistically significant ($P > .025$). Note that these occurrences would not be false positives because the null hypothesis that the survival curves are identical is not true. However, if one believed that it would be misleading to the clinical community to see only the first 6 years of the curves in Figure 1, then releasing the results early would be a mistake regardless of the statistical significance of the early results. In practice, one will unlikely know beforehand whether the curves will come back together as in Figure 1 or keep separating. This suggests that interim monitoring is appropriate, but additional follow-up after the early release of extremely positive results is advisable.

Empirical evidence of crossing hazards is suggested by Figures 2 and 3, which display the OS curves for EST-3189 and the complete continuous remission curves for POG-9006, respectively. For EST-3189, the *P* value went from .0025 at the time of interim monitoring boundary crossing to .10 with 2 years of additional follow-up. For POG-9006, the *P* value went from .0016 at the time of interim monitoring boundary crossing to .22 with 4 years of additional follow-up.

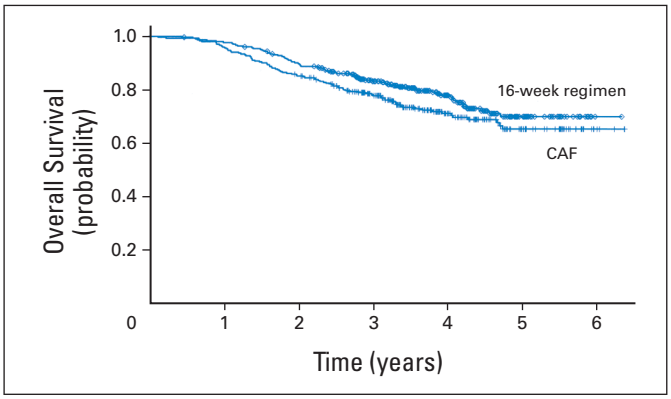

Fig 2. Overall survival curves from EST-3189 (control treatment = cyclophosphamide, doxorubicin, and fluorouracil [CAF]; experimental treatment = 16-week regimen; $P = .10$). (Reprinted from Fig 2 with permission.[14])

Type I Errors

Although RCTs are designed to answer definitively a treatment question, they are not perfect. In particular, they will infrequently lead to a rejection of the null hypothesis when it is true (a type I error) or the nonrejection of the null hypothesis when the treatments are truly different (a type II error). A design parameter for RCTs is the type I error rate, which is frequently set at 0.05. This means that if the null hypothesis is true, then there is, at most, a 5% chance that the trial will result in a statistically significant outcome. It is important to note that, in a properly designed trial, the type I error rate encompasses both type I errors that occur when the trial is stopped early for positive results as well as type I errors that occur with a positive conclusion at the regularly scheduled trial end. Therefore, there is not an excess of type I errors as a result of the possibility of early stopping with appropriately designed interim monitoring boundaries.

An alternative way to consider type I errors vis-à-vis concerns about early stopping is to calculate the probability that a positive conclusion is a false positive, given that the trial stopped early. A standard application of Bayes' theorem allows this calculation as a function of the prior probability that the null hypothesis is true.[64] Such calculations show that, for a positive trial, the probability that the trial is a false positive is lower if the trial crossed an interim monitoring boundary than if it did not. As a simple example, consider a trial designed with 90% power for a specified alternative, with one-sided type I error of 0.025, and where the true treatment effect is null 80% of the time and equal to the specified alternative 20% of the time. Without the possibility of early stopping, the probability that a trial with a positive outcome (one-sided $P < .025$) is a false positive is 10%. If an O'Brien-Fleming interim monitoring boundary[65] is used with two equally spaced interim looks, then the probability that a trial that crosses this boundary at the first interim look (33% information) is a false positive is 1.2%, and the probability that a trial that first crosses the boundary at the second interim look (67% information) is a false positive is 4.4%; the overall false-positive rate remains at 10%.

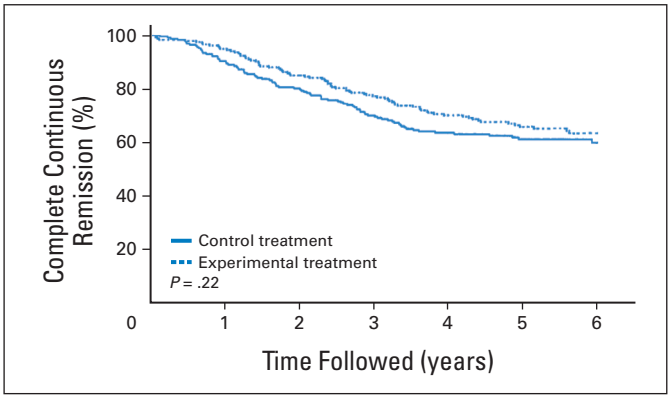

Fig 3. Complete continuous remission curves from Pediatric Oncology Group trial 9006. (Adapted from Fig 2[12] by permission from Macmillan Publishers Ltd.)

Biased or Implausible Positive Interim Results

It has been noted[66,67] that a treatment effect observed for a trial that stops early for positive results will be, on average, higher than the true treatment effect (ie, is biased upward). Some[2,3,5] use this to argue against stopping trials early for positive effects. However, it is also true that the observed treatment effect for a trial that concludes at its regularly scheduled end with a significantly positive result will be biased upward (although not as high as one that has stopped early).[68] In particular, for interim analyses occurring with half or more of the total planned events, the upward bias as a result of early stopping is comparable to the upward bias seen in similarly positive trials not stopped early.[69] It is important to note that even though the treatment effect is biased upward when estimated when a trial stops early for positive results, there is only a small probability (the type I error) that the treatment effect is not positive. Therefore, concerns about treating future patients with the best treatment may outweigh concerns about not knowing exactly how much better the better treatment is. The empirical data in Table 2 suggest that the potential bias as a result of early stopping is not a major problem.

It has been suggested[70] that if the magnitude of the treatment effect at an interim monitoring look is implausible, then one should not stop the trial at this point (implying that one should stop for a smaller observed effect). However, not stopping a trial for extremely positive results but stopping it for less extreme positive results runs counter to both common sense and statistical thinking[71]; see Clayton and Wheatley[72] for an alternative point of view. Whether or not a trial is stopped early, if one has prior information about the magnitude of the treatment effect, then a Bayesian analysis[73] may be useful in providing an attenuated estimator of an extremely positive treatment effect.

DISCUSSION

The vast majority of NCI Cooperative Group phase III trials that crossed an interim monitoring boundary for positive results led to the early release of treatment effect data to the public that, in retrospect, was appropriate and beneficial. Concerns about excess false positives as a result of the early stopping are not supported by statistical theory or the empirical evidence presented here. Concerns about biased treatment effects as a result of the early stopping are statistically valid but may not be practically important; the bias may not be much larger than would be seen for a positive trial not stopped early, and releasing information early about an effective treatment may be more important than knowing the exact magnitude of the benefit. However, concerns about early stopping/release limiting the ability to estimate long-term survival curves (and potentially identify crossing hazards)

or to estimate OS curves (when the stopping is based on a non-OS end point) are statistically valid and practically important. An important consideration in this situation is whether the survival curves can be accurately estimated with additional follow-up after the early stopping/release. If the accrual was not complete at the time of early stopping or many patients could be expected to cross over to the experimental treatment when the positive results are released, then it may be impossible even with additional follow-up to estimate what the survival curves would have looked like if there had been no early stopping/release. In this situation, the interim monitoring plan could be conservative during accrual if the monitoring end point is not OS or there is strong interest in the long-term survival curves.

The NCI Cooperative Group trials that we have considered had well-designed interim monitoring plans. The choice of end point and monitoring plan needs to be carefully considered before a trial starts; trial investigators should be comfortable with the predictable stopping and not stopping decisions that will occur under different accruing data scenarios. The ability to stop a trial and release positive data early is an important component of phase III trial design, allowing the public to benefit as soon as possible from the study conclusions.

Note Added in Proof

After this article was accepted for publication, another trial came to our attention that was stopped early for positive results.[74] The trial was not identified in our search because the abstract reporting the initial results[75] made no mention that the trial was stopped early.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The author(s) indicated no potential conflicts of interest.

AUTHOR CONTRIBUTIONS

Conception and design: Edward L. Korn, Boris Freidlin, Margaret Mooney
Collection and assembly of data: Edward L. Korn, Boris Freidlin
Data analysis and interpretation: Edward L. Korn, Boris Freidlin
Manuscript writing: Edward L. Korn, Boris Freidlin, Margaret Mooney
Final approval of manuscript: Edward L. Korn, Boris Freidlin, Margaret Mooney

REFERENCES

1. Montori VM, Devereaux PH, Adhikari NKJ, et al: Randomized trials stopped early for benefit: A systematic review. JAMA 294:2203-2209, 2005

2. Mueller PS, Montori VM, Bassler D, et al: Ethical issues in stopping randomized trials early because of apparent benefit. Ann Intern Med 146: 878-881, 2007

3. Bassler D, Montori VM, Briel M, et al: Early stopping of randomized clinical trials for overt efficacy is problematic. J Clin Epidemiol 61:241-246, 2008

4. Wilcox RA, Djulbegovic B, Guyatt GH, et al: Randomized trials in oncology stopped early for benefit. J Clin Oncol 26:18-19, 2008

5. Trotta F, Apolone G, Garattini S, et al: Stopping a trial early in oncology: For patients or for industry? Ann Oncol 19:1347-1353, 2008

6. Pater JL, Goss P, Meyer R: Stopping trials for benefit can (sometimes) benefit patients. J Clin Oncol 16:2787-2788, 2008

7. Moertel CG, Fleming TR, Macdonald JS, et al: Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. N Engl J Med 322:352-358, 1990

8. Moertel CG, Fleming TR, Macdonald JS, et al: Fluorouracil plus levamisole as effective adjuvant therapy after resection of stage III colon carcinoma: A final report. Ann Intern Med 122:321-326, 1995

9. 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 326:1593-1598, 1992

10. 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). JAMA 281:1623-1627, 1999

11. Lauer SJ, Toledano S, Winick N, et al: A comparison of early intensive methotrexate/mercaptopurine (MTX/MP) vs early intensive alternating chemotherapy for high-risk acute lymphoblastic leukemia (HR-ALL): A Pediatric Oncology Group (POG) randomized phase III study. Proc Am Soc Clin Oncol 14:342, 1995 (abstr A1030)

12. Lauer SJ, Shuster JJ, Mahoney DH, et al: A comparison of early intensive methotrexate/mercaptopurine with early intensive alternating combination chemotherapy for high-risk B-precursor acute lymphoblastic leukemia: A Pediatric Oncology Group phase III randomized trial. Leukemia 15:1038-1045, 2001

13. Fetting J, Gray R, Abeloff M, et al: CAF vs a 16-week multidrug regimen as adjuvant therapy for receptor-negative, node-positive breast cancer: An intergroup study. Proc Am Soc Clin Oncol 14:96, 1995 (abstr A83)

14. Fetting JH, Gray R, Fairclough DL, et al: Sixteen-week multidrug regimen versus cyclophosphamide, doxorubicin, and fluorouracil as adjuvant therapy for node-positive, receptor-negative breast cancer: An Intergroup study. J Clin Oncol 16:2382-2391, 1998

15. Rai KR, Peterson B, Kolitz J, et al: Fludarabine induces a high complete remission rate in previously untreated patients with active chronic lymphocytic leukemia (CLL): A randomized inter-group study. Blood 86:A607, 1995 (abstr 2414)

16. Rai KR, Peterson BL, Appelbaum FR, et al: Fludarabine compared with chlorambucil as primary therapy for chronic lymphocytic leukemia. N Engl J Med 343:1750-1757, 2000

17. Al-Sarraf M, LeBlanc M, Giri PG, et al: Superiority of chemoradiotherapy (CT-RT) vs. radiotherapy (RT) in patients (pts) with locally advanced nasopharyngeal cancer (NPC): Preliminary results of intergroup (0099) (SWOG 8892, RTOG 8817, ECOG 2388) randomized study. Proc Am Soc Clin Oncol 15:313a, 1996 (abstr A882)

18. Al-Sarraf M, LeBlanc M, Giri PG, et al: Chemoradiotherapy versus radiotherapy in patients with advanced nasopharyngeal cancer: Phase III randomized Intergroup study 0099. J Clin Oncol 16:1310-1317, 1998

19. Al-Sarraf M, Le Blanc M, Giri P, et al: Superiority of five year survival with chemo-radiotherapy (CT-RT) vs radiotherapy in patients with locally advanced nasopharyngeal cancer (NPC): Intergroup (0099) (SWOG 8892, RTOG 8817, ECOG 2388) phase III study—Final report. Proc Am Soc Clin Oncol 20:227a, 2001 (abstr A905)

20. Tallman MS, Andersen JW, Schiffer CA, et al: All-trans-retinoic acid in acute promyelocytic leukemia. N Engl J Med 337:1021-1028, 1997

21. Nachman JB, Sather HN, Sensel MG, et al: Augmented post-induction therapy for children with high-risk acute lymphoblastic leukemia and a slow response to initial therapy. N Engl J Med 338:1663-1671, 1998

22. Albain K, Green S, Osborne K, et al: Tamoxifen (T) versus cyclophosphamide, Adriamycin and 5-FU plus either concurrent or sequential T in postmenopausal, receptor(+), node(+) breast cancer: A Southwest Oncology Group phase III intergroup trial (SWOG-8814, INT-0100). Proc Am Soc Clin Oncol 16:128a, 1997 (abstr A450)

23. Albain K, Barlow W, O'Malley F, et al: Concurrent (CAFT) versus sequential (CAF-T) chemohormonal therapy (cyclophosphamide, doxorubicin, 5-fluorouracil, tamoxifen) versus T alone for postmenopausal, node-positive, estrogen (ER) and/or progesterone (PgR) receptor-positive breast cancer: Mature outcomes and new biologic correlates on phase III intergroup trial 0100 (SWOG-8814). Breast Cancer Res Treat 90:195, 2005 (abstr A37)

24. Peters WA, Liu PY, Barrett R, et al: Cisplatin, 5-fluorouracil plus radiation therapy are superior to radiation therapy as adjunctive therapy in high-risk, early stage carcinoma of the cervix after radical hysterectomy and pelvic lymphadenectomy: Report of a phase III intergroup study. Gynecol Oncol 72:443, 1999 (abstr A1)

25. Peters WA 3rd, Liu PY, Barrett RJ 2nd, 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 18: 1606-1613, 2000

26. Henderson IC, Berry D, Demetri G, et al: Improved disease-free (DFS) and overall survival (OS) from the addition of sequential paclitaxel (T) but not from the escalation of doxorubicin (A) dose level in the adjuvant chemotherapy of patients (PTS) with

node-positive primary breast cancer. Proc Am Soc Clin Oncol 17:101a, 1998 (abstr 390)

27. Henderson IC, Berry DA, Demetri GD, et al: Improved outcomes from adding sequential paclitaxel but not from escalating doxorubicin dose in an adjuvant chemotherapy regimen for patients with node-positive primary breast cancer. J Clin Oncol 21:976-983, 2003

28. George SI, Green MR: Controversies in the early reporting of a clinical trial in early breast cancer, in DeMets DL, Furberg CD, Friedman LM (eds): Data Monitoring in Clinical Trials. New York, NY, Springer, 2006, pp 346-359

29. Morris M, Eifel PJ, Lu J, et al: Pelvic radiation with concurrent chemotherapy compared with pelvic and para-aortic radiation for high-risk cervical cancer. N Engl J Med 340:1137-1143, 1999

30. Eifel PJ, Winter K, Morris M, et al: Pelvic irradiation with concurrent chemotherapy versus pelvic and para-aortic irradiation for high-risk cervical cancer: An update of Radiation Therapy Oncology Group trial (RTOG) 90-01. J Clin Oncol 22:872-880, 2004

31. Nachman JB, Sposto R, Herzog P, et al: Randomized comparison of low-dose involved-field radiotherapy and no radiotherapy for children with Hodgkin's disease who achieve a complete response to chemotherapy. J Clin Oncol 20:3765-3771, 2002

32. Press OW, LeBlanc M, Lichter A, et al: A phase III randomized intergroup trial of subtotal lymphoid irradiation (STLI) versus doxorubicin, vinblastine, and STLI for stage IA-IIA Hodgkin's disease (SWOG 9133, CALGB 9391). Blood 96:575a, 2000 (abstr A2471)

33. Press OW, LeBlanc M, Lichter AS, et al: Phase III randomized intergroup trial of subtotal lymphoid irradiation versus doxorubicin, vinblastine, and subtotal lymphoid irradiation for stage IA to IIA Hodgkin's disease. J Clin Oncol 19:4238-4244, 2001

34. Roach M III, 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 21: 1904-1911, 2003

35. Lawton CA, DeSilvio M, Roach M 3rd, 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 69:646-655, 2007

36. Roach M III, Lu JD, Lawton C, et al: A phase III trial comparing whole-pelvic (WP) to prostate only (PO) radiotherapy and neoadjuvant to adjuvant total androgen suppression (TAS): Preliminary analysis of RTOG 9413. Int J Radiat Oncol Biol Phys 51:3, 2001 (suppl; abstr plenary 5)

37. 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 21:2460-2465, 2003

38. Markman M, Liu P, Wilczynski S, et al: Survival (S) of ovarian cancer (OC) patients (pts) treated on SWOG9701/GOG178: 12 versus (v) 3 cycles (C) of monthly single-agent paclitaxel (PAC) following attainment of a clinically-defined complete response (CR) to platinum (PLAT)/PAC. J Clin Oncol 24:257s, 2006 (suppl; abstr 5005)

39. Goldberg RM, Morton RF, Sargent DJ, et al: N9741: Oxaliplatin (oxal) or CPT-11 + 5-fluorouracil (5FU)/leucovorin (LV) or oxal + CPT-11 in advanced colorectal cancer (CRC)—Initial toxicity and response data from a GI Intergroup study. Proc Am Soc Clin Oncol 21:128a, 2002 (abstr 511)

40. Goldberg RM, Sargent DJ, Morton RF, et al: A randomized controlled trial of fluorouracil plus leucovorin, irinotecan, and oxaliplatin combinations in patients with previously untreated metastatic colorectal cancer. J Clin Oncol 22:23-30, 2004

41. 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 349:1793-1802, 2003

42. Goss PE, Ingle JN, Martino S, et al: Randomized trial of letrozole following tamoxifen as extended adjuvant therapy in receptor-positive breast cancer: Updated findings from NCIC CTG MA. 17. J Natl Cancer Inst 97:1262-1271, 2005

43. Hochster HS, Weller E, Ryan T, et al: Results of E1496: A phase III trial of CVP with or without maintenance rituximab in advanced indolent lymphoma (NHL). J Clin Oncol 22:558s, 2004 (suppl 14; abstr 6502)

44. Hochster HS, Weller E, Gascoyne RD, et al: Maintenance rituximab after CVP results in superior clinical outcome in advanced follicular lymphoma (FL): Results of the E1496 phase III trial from the Eastern Cooperative Oncology Group and the Cancer and Leukemia Group B. Blood 106:106a, 2005 (abstr 349)

45. Rajkumar SV, Blood E, Vesole DH, et al: A randomised phase III trial of thalidomide plus dexamethasone versus dexamethasone in newly diagnosed multiple myeloma (E1A00): A trial coordinated by the Eastern Cooperative Oncology Group. J Clin Oncol 22:560s, 2004 (suppl; abstr 6508)

46. Rajkumar SV, Blood E, Vesole D, et al: Phase III clinical trial of thalidomide plus dexamethasone compared with dexamethasone alone in newly diagnosed multiple myeloma: A clinical trial coordinated by the Eastern Cooperative Oncology Group. J Clin Oncol 24:431-436, 2006

47. Seibel NL, Steinherz PG, Sather HN, et al: Early treatment intensification improves outcome in children and adolescents with acute lymphoblastic leukemia (ALL) presenting with unfavorable features who show a rapid early response (RER) to induction chemotherapy: A report of CCG1961. Blood 102: 224A-225A, 2003 (abstr 787)

48. Seibel NL, Steinherz PG, Sather HN, et al: Early postinduction intensification therapy improves survival for children and adolescents with high-risk acute lymphoblastic leukemia: A report from the Children's Oncology Group. Blood 111:2548-2555, 2008

49. Giantonio BJ, Catalano PJ, Meropol NJ, et al: High-dose bevacizumab improves survival when combined with FOLFOX4 in previously treated advanced colorectal cancer: Results from the Eastern Cooperative Oncology Group (ECOG) study E3200. J Clin Oncol 23:1s, 2005 (suppl; abstr 2)

50. Giantonio BJ: ECOG-E3200: FOLFOX4 with bevacizumab versus FOLFOX4 versus bevacizumab in patients with previously treated advanced colorectal cancer. Colorectal Cancer Update 4:18-21, 2005

51. Giantonio BJ, Catalano PJ, Meropol NJ, et al: Bevacizumab in combination with oxaliplatin, fluorouracil, and leucovorin (FOLFOX4) for previously treated metastatic colorectal cancer: Results from the Eastern Cooperative Oncology Group Study E3200. J Clin Oncol 25:1539-1544, 2007

52. Flinn IW, Kumm E, Grever MR, et al: Fludarabine and cyclophosphamide produces a higher complete response rate and more durable remissions than fludarabine in patients with previously untreated CLL: Intergroup trial E2997. Blood 104:139a, 2004 (abstr 475)

53. Flinn IW, Neuberg DS, Grever MR, et al: Phase III trial of fludarabine plus cyclophosphamide compared with fludarabine for patients with previously untreated chronic lymphocytic leukemia: US Intergroup Trial E2997. J Clin Oncol 25:793-798, 2007

54. Romond EH, Perez EA, Bryant J, et al: Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. N Engl J Med 353: 1673-1684, 2005

55. Perez E, Romond E, Suman V, et al: Updated results of the combined analysis of NCCTG N9831 and NSABP B-31 adjuvant chemotherapy with/without trastuzumab in patients with HER2-positive breast cancer. J Clin Oncol 25:6s, 2007 (suppl; abstr 512)

56. Sandler A, Gray R, Perry MC, et al: Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer. N Engl J Med 355:2542-2550, 2006

57. Miller KD, Wang M, Gralow J, et al: A randomized phase III trial of paclitaxel versus paclitaxel plus bevacizumab as first-line therapy for locally recurrent or metastatic breast cancer: A trial coordinated by the Eastern Cooperative Oncology Group (E2100). Breast Cancer Res Treat 94:S6, 2005 (suppl 1; abstr 3)

58. Miller K, Wang M, Gralow J, et al: Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer. N Engl J Med 357:2666-2676, 2007

59. Burnell MJ, Levine MN, Chapman JA, et al: A phase III adjuvant trial of sequenced EC + filgrastim + epoetin-alpha followed by paclitaxel compared to sequenced AC followed by paclitaxel compared to CEF in women with node-positive or high-risk node-negative breast cancer (NCIC CTG MA. 21). J Clin Oncol 25:15s, 2007 (suppl; abstr 550)

60. Burnell M, Levine M, Chapman JA, et al: A randomized trial of CEF versus dose dense EC followed by paclitaxel versus AC followed by paclitaxel in women with node positive or high risk node negative breast cancer, NCIC CTG MA. 21: Results of an interim analysis. 29th Annual San Antonio Breast Cancer Symposium, San Antonio, TX, December 14-17, 2006 (abstr A53)

61. Cannistra SA: The ethics of early stopping rules: Who is protecting whom? J Clin Oncol 22: 1542-1545, 2005

62. Pater J, Goss P, Ingel J, et al: The ethics of early stopping rules. J Clin Oncol 23:2862-2863, 2005

63. Bukowski RM, Eisen T, Szczylik C, et al: Final results of the randomized phase III trial of sorafenib in advanced renal cell carcinoma: Survival and biomarker analysis. J Clin Oncol 25:240s, 2007 (suppl; abstr 5023)

64. Peto R, Pike MC, Armitage P, et al: Design and analysis of randomized clinical trials requiring prolonged observation of each patient: I. Introduction and design. Br J Cancer 34:535-612, 1976

65. O'Brien PC, Fleming TR: A multiple testing procedure for clinical trials. Biometrics 35:549-556, 1979

66. Pocock SJ, Hughes MD: Practical problems in interim analyses, with particular regard to estimation. Control Clin Trials 10:209S-221S, 1989

67. Fan X, DeMets DL, Lan KKG: Conditional bias of point estimates following a group sequential test. J Biopharm Stat 14:505-530, 2004

68. Goodman SN: Stopping at nothing? Some dilemmas of data monitoring in clinical trials. Ann Intern Med 146:882-887, 2007

69. Freidlin B, Korn EL: Stopping clinical trials early for benefit: Implact on estimation. Clin Trials (in press)

70. Wheatley K, Clayton D: Be skeptical about unexpected large apparent treatment effects: The case of an MRC AML12 randomization. Control Clin Trials 24:66-70, 2003

71. Korn EL, Freidlin B, George SL: Data monitoring and large apparent treatment effects. Control Clin Trials 25:67-69, 2004

72. Clayton D, Wheatley K: Reply to the letter of Korn, et al. Control Clin Trials 25:71-72, 2004

73. Spiegelhalter DJ, Freedman LS, Parmar MKB: Bayesian approaches to randomized trials. J R Stat Soc A 157:357-387, 1994

74. Strauss GM, Herndon JE II, Maddaus MA, et al: Adjuvanct paclitaxel plus carboplatin compared with observation in stage 1B non–small-cell lung cancer: CALGB 9633 with the Cancer and Leukemia Group B, Radiation Therapy Oncology Group, and North Central Cancer Treatment Group Study Groups. J Clin Oncol 31:5043-5051, 2008

75. Strauss GM, Herndon J, Maddaus MA, et al: Randomized clinical trial of adjuvant chemotherapy with paclitaxel and carboplatin following resection in stage 1B non–small cell lung cancer (NSCLC): Report of Cancer and Leukemia Group B (CALGB) protocol 9633. J Clin Oncol 22:621s, 2004 (suppl; abstr 7019)

Acknowledgment

We thank W. Barlow, M. Devidas, R.J. Gray, P.Y. Liu, M. LeBlanc, B. Peterson, N.L. Seibel, R. Sposto, and K. Winter for providing some statistical details concerning previous trial analyses and R. Gore-Langton for his help in locating publications.