Stopping a trial early in oncology: for patients or for industry?

F. Trotta¹, G. Apolone², S. Garattini² & G. Tafuri¹,3*
¹Italian Medicines Agency (AIFA), Rome; ²Mario Negri Institute for Pharmacological Research, Milan, Italy; ³Utrecht University, Utrecht Institute for Pharmaceutical Sciences, Utrecht, The Netherlands

Received 18 December 2007; revised 25 January 2008; accepted 28 January 2008

**Background:** The aim of this study is to assess the use of interim analyses in randomised controlled trials (RCTs) testing new anticancer drugs, focussing on oncological clinical trials stopped early for benefit.

**Materials and methods:** All published clinical trials stopped early for benefit and published in the last 11 years, regarding anticancer drugs and containing an interim analysis, were assessed.

**Results:** Twenty-five RCTs were analysed. The evaluation of efficacy was protocol planned through time-related primary end points, >40% of them overall survival. In 95% of studies, at the interim analysis, efficacy was evaluated using the same end point as planned for the final analysis. As a consequence of early stopping after the interim analysis, ~3300 patients/events across all studies were spared. More than 78% of the RCTs published in the last 3 years were used for registration purposes.

**Conclusion:** Though criticism of the poor quality of oncological trials seems out of place, unfortunately early termination raises new concerns. The relation between sparing patients and saving time and trial costs indicates that there is a market-driven intent. We believe that only untruncated trials can provide a full level of evidence which can be translated into clinical practice without further confirmative trials.

**Key words:** anticancer drugs, EMEA, FDA, end point, interim analysis, RCT

---

**background**

European legislation in pharmaceuticals has been recently revised. In line with the path of the United States Food and Drug Administration (FDA), new procedures for granting marketing authorisation now include accelerated and conditional approvals, leading to quicker access of new drugs to patients. In this evolving scenario, guidelines on the evaluation of medicinal products are subject to continuous revision. This is especially the case for anticancer drugs, for which important changes have been made in terms of trial design and conduct [1]. These factors have simplified and shortened the process of development of a new drug, particularly in oncology.

Previous analyses of new anticancer compounds, approved by both the European Medicines Agency (EMEA) and the FDA, highlighted methodological concerns in terms of lack of comparative trials, use of surrogate end points, lack of evidence for establishing the added value, and lack of blinding/masking [2–5]. In addition, a recent systematic review covering different therapeutic areas found that the number of randomised trials stopped early for benefit had more than doubled since 1990 [6].

Interim analyses pose the ethical dilemma of safeguarding the interests of patients enrolled in clinical trials while also protecting society from overzealous premature claims of treatment benefit. Trials stopped early because of harm (toxicity) or futility tend to result in prompt discontinuation of useless or potentially harmful interventions. In contrast, trials stopped early for benefit may result in the quick identification, approval, and dissemination of promising new treatments.

Given the serious and life-threatening nature of cancer and patients’ expectations, quicker clinical drug development is required by both patients and clinicians, but this may lead to an unclear and poorly defined benefit/risk balance of new drugs.

**objective**

The aim of this study is to assess the use of interim analyses in randomised controlled trials (RCTs) testing new anticancer drugs, focussing on oncological clinical trials stopped early for benefit. A second aim is to estimate how often trials prematurely stopped as a result of an interim analysis are used for registration purposes. Our study presents an updated overview of this growing phenomenon in the specific field of oncology, which is subject to continuous change. The analysis...
focussed on trials that were halted after an interim analysis found the treatment carried out better than the control arm.

**materials and methods**

All clinical trials published from January 1997 to October 2007, regarding anticancer drugs and containing an interim analysis, were retrieved through Medline. The following strategy was adopted: publications containing the words 'interim' and 'analysis', and limited to humans, clinical trials, cancer, and English language, were searched. A total of 231 reports were found.

In order to test the sensitivity of the research methods, we did an extra search for articles, that might have been missed in the first search, published in the three main peer-reviewed journals (The Lancet, The New England Journal of Medicine, and The Journal of Clinical Oncology) from October 2006 to October 2007; this produced two more reports. These three journals were chosen on the basis of how frequently they had reported the articles retrieved through the Medline search.

To increase the specificity, articles were initially screened on the basis of the abstract. Of the 233, 140 reports were excluded as not relevant according to predefined inclusion and exclusion criteria (see Figure 1). In particular, phase I trials, trials testing growth factors, and those based solely on surgery or radiotherapy were considered not pertinent, so were excluded. Furthermore, studies comparing different dose regimens and schedules of the same drugs and studies on the basis of palliative/supportive therapies (e.g. antiemetics) were excluded. Study protocols were also excluded.

Only papers on the basis of trials of anticancer medicinal products and containing an interim analysis were initially considered eligible for analysis (93 papers describing 93 trials). Out of these 93 papers, 65 were subsequently excluded for the following reasons: 4 trials were stopped after an interim analysis because of harm (toxicity) and 28 because of futility (lack of efficacy). Another 33 papers were excluded because the trials were not actually stopped after the interim analysis and were thus considered ongoing.

Twenty-eight papers met the inclusion criteria, i.e. clinical trials testing anticancer medicinal products truncated for benefit after an interim analysis. However, one was unretrievable, and in two separate cases two papers reported the same study in two different journals. In both cases, only the paper published earlier was included. Following these corrections, a final sample of 25 papers describing 25 trials was obtained.

For the purpose of this analysis, the following parameters were assessed: disease, study duration, date of publication, presence of a ‘Data and Safety Monitoring Committee’ (DSMC), type of end point(s), sample size, rationale for interim analysis and type of analysis carried out, consequences of the interim analysis on the RCT and on the patients, and characteristics of the control group. We reported both the primary end point planned for final efficacy analysis and the end point used for the interim analysis. The same was done for sample size. In a few cases, certain study characteristics were not reported in the data acquisition form because they were not specified in the analysed articles. The investigators defined a priori a common data acquisition form to be completed. FT and GT independently evaluated all the selected papers and filled the respective forms. The results were then cross-checked, leading to a joint document. In the case of disagreement, the final decision was taken through a consensus process reached following a discussion.

### results

Of the 93 papers, initially selected as having been stopped after an interim analysis, 28 (30%) were stopped early for benefit, 28 (30%) for futility, and 4 (4%) for harm.

As described above, 25 of 28 papers were actually included in the analysis (see Figure 1). All 25 were RCTs, on a variety of different cancers (Table 1). In 16 trials, the control arm used an active comparator and in four used a placebo, while in five no treatment was given. In no case was information provided about trial design in terms of superiority, non-inferiority, or equivalence.

More than half of the selected trials (56%) were published in the last 3 years (2005–2007), 11 of them (79%) were used to support an application for marketing authorisation at the EMEA and FDA (Table 2).

The evaluation of efficacy was protocol planned through time-related primary end points, >40% of them overall survival (10 of 23, as information was unavailable in two cases). In two cases, publications lacked a clear definition of the primary end points, for both the final and the interim analyses. In 95% of studies (22 of 23), at the interim analysis, efficacy was evaluated using the same end point as planned for the final analysis. There was no DSMC in 24% (6 of 25) of the studies.

All RCTs reported consequences after the interim analysis. Those fell into three groups: cross-over to the treatment group, stopping enrolment, and disclosure of results (Table 2).

The criteria for planning an interim analysis were based either on a cut-off date (3 of 24, as in one case information was missing) or on the number of observed events (12 of 24) or
<table>
<thead>
<tr>
<th>Treatment</th>
<th>Disease</th>
<th>Trial starting date versus end of the enrolment</th>
<th>DSMC</th>
<th>Primary end point versus end point used in interim analysis</th>
<th>Planned versus interim sample size</th>
<th>Rationale for planning interim analysis</th>
<th>Interim analysis consequences</th>
<th>Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Melphalan + prednisone + thalidomide&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Multiple Myeloma</td>
<td>May 2000 versus August 2005 versus October 2007</td>
<td>Y</td>
<td>OS versus OS</td>
<td>500 versus 447 patients</td>
<td>No. of enrolled patients</td>
<td>Stop enrolment</td>
<td>A</td>
</tr>
<tr>
<td>Temsirolimus&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Advanced renal cell carcinoma</td>
<td>July 2003 versus April 2005 versus May 2007</td>
<td>Y</td>
<td>OS versus OS</td>
<td>600 versus 446 events</td>
<td>No. of events</td>
<td>Disclosure of results</td>
<td>A</td>
</tr>
<tr>
<td>Sunitinib&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Metastatic renal cell carcinoma</td>
<td>August 2004 versus October 2005 versus January 2007</td>
<td>Y</td>
<td>PFS versus PFS</td>
<td>471 events versus NA</td>
<td>Cut-off date</td>
<td>Crossover to treatment group</td>
<td>A</td>
</tr>
<tr>
<td>Sorafenib&lt;sup&gt;*&lt;/sup&gt;</td>
<td>Advanced clear-cell renal cell carcinoma</td>
<td>November 2003 versus March 2005 versus January 2007</td>
<td>Y</td>
<td>OS versus PFS</td>
<td>540 versus 363 events</td>
<td>No. of events</td>
<td>Crossover to treatment group</td>
<td>P</td>
</tr>
<tr>
<td>Paclitaxel + carboplatin + bevacizumab&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Non-small-cell lung cancer</td>
<td>July 2001 versus April 2004 versus December 2006</td>
<td>Y</td>
<td>OS versus OS</td>
<td>650 versus 455 events</td>
<td>No. of events</td>
<td>Disclosure of results</td>
<td>A</td>
</tr>
<tr>
<td>Lapatinib + capecitabine&lt;sup&gt;a&lt;/sup&gt;</td>
<td>HER-2-positive metastatic breast cancer</td>
<td>March 2004 versus NA versus December 2006</td>
<td>Y</td>
<td>TtP versus TtP</td>
<td>266 versus 114 events</td>
<td>No. of events</td>
<td>Crossover to treatment group + stop enrolment</td>
<td>A</td>
</tr>
<tr>
<td>Sunitinib&lt;sup&gt;*&lt;/sup&gt;</td>
<td>Advanced gastrointestinal stromal tumour</td>
<td>December 2003 versus January 2005 versus October 2006</td>
<td>Y</td>
<td>TtP versus TtP</td>
<td>281 versus 149 events</td>
<td>No. of events</td>
<td>Crossover to treatment group</td>
<td>P</td>
</tr>
<tr>
<td>FU + leucovorina + oxaliplatin (FOLFOX4)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Metastatic colorectal cancer</td>
<td>April 2001 versus April 2002 versus July 2006</td>
<td>Y</td>
<td>TtP versus TtP</td>
<td>550 versus 305 patients</td>
<td>No. of events</td>
<td>Stop enrolment</td>
<td>A</td>
</tr>
<tr>
<td>Uracil + tegafur</td>
<td>Stage III rectal cancer</td>
<td>October 1996 versus April 2001 versus April 2006</td>
<td>Y</td>
<td>RFS versus RFS and OS</td>
<td>400 versus 274 patients</td>
<td>Cut-off date</td>
<td>Disclosure of results</td>
<td>NT</td>
</tr>
<tr>
<td>Carboplatin</td>
<td>Epithelial ovarian cancer</td>
<td>November 1994 versus July 1998 versus January 2006</td>
<td>Y</td>
<td>TtP versus TtP</td>
<td>190 versus 120 events</td>
<td>No. of events</td>
<td>Stop enrolment</td>
<td>A</td>
</tr>
<tr>
<td>Methotrexate, vinblastine, doxorubicin/epirubicin, cisplatin</td>
<td>Advanced bladder cancer</td>
<td>May 1987 versus December 1990 versus January 2006</td>
<td>N</td>
<td>PFS versus PFS</td>
<td>NA</td>
<td>NA</td>
<td>Stop enrolment</td>
<td>NT</td>
</tr>
<tr>
<td>Trastuzumab + adjuvant chemotherapy&lt;sup&gt;a&lt;/sup&gt;</td>
<td>HER-2 positive breast cancer</td>
<td>May 2003 versus November 2004 versus October 2005</td>
<td>Y</td>
<td>DFS versus DFS</td>
<td>710 versus 394 events</td>
<td>No. of events</td>
<td>Stop enrolment + disclosure of results</td>
<td>A</td>
</tr>
<tr>
<td>Trastuzumab (1-year arm + 2-year arm)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>HER-2-positive early-stage invasive breast cancer</td>
<td>December 2001 versus March 2005 versus October 2005</td>
<td>Y</td>
<td>DFS versus DFS</td>
<td>951 versus 347 events</td>
<td>No. of events</td>
<td>Disclosure of results</td>
<td>NT</td>
</tr>
<tr>
<td>Treatment</td>
<td>Disease</td>
<td>Trial starting date versus end of the enrolment versus date of publication</td>
<td>DSMC Primary end point versus end point used in interim analysis</td>
<td>Planned versus interim sample size</td>
<td>Rationale for planning interim analysis</td>
<td>Interim analysis consequences</td>
<td>Control group</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>--------</td>
<td>--------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------</td>
<td>-----------------------------------</td>
<td>--------------------------------------</td>
<td>-------------------------------</td>
<td>--------------</td>
<td></td>
</tr>
<tr>
<td>Bevacizumab + (irinotecan, FU, leucovorin) and bevacizumab + (FU, leucovorin)*</td>
<td>Metastatic colorectal cancer</td>
<td>NA versus NA versus May 2005</td>
<td>Y OS versus OS</td>
<td>After 313 patients enrolled</td>
<td>No. of enrolled patients</td>
<td>Stop enrolment</td>
<td>A+ P</td>
<td></td>
</tr>
<tr>
<td>Letrozole*</td>
<td>Adjuvant therapy in receptor-positive breast cancer</td>
<td>August 1998 versus September 2002 versus November 2003</td>
<td>Y DFS versus DFS</td>
<td>515 versus 171 events</td>
<td>No. of events</td>
<td>Disclosure of results</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>Gemcitabine</td>
<td>Advanced or metastatic adenocarcinoma of the pancreas</td>
<td>December 1997 versus July 1999 versus September 2003</td>
<td>Y OS versus OS and PFS</td>
<td>350 patients versus 140 events</td>
<td>No. of events</td>
<td>Stop enrolment + crossover to treatment group</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>Idoxifene</td>
<td>Postmenopausal metastatic breast cancer</td>
<td>December 1996 versus May 1999 versus February 2003</td>
<td>N RR, TtP versus RR, TtP</td>
<td>440 versus 321 patients</td>
<td>No. of enrolled patients</td>
<td>Trial stopped for economic consideration</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>ChlVPP/EVA hybrid</td>
<td>Hodgkin’s disease</td>
<td>September 1992 versus September 1996 versus July 2002</td>
<td>N NA</td>
<td>80 versus 60 events</td>
<td>No. of events</td>
<td>Stop enrolment</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>Irinotecan + cisplatin</td>
<td>Metastatic small-cell lung cancer</td>
<td>November 1995 versus November 1998 versus January 2002</td>
<td>Y OS versus OS</td>
<td>230 versus 230 patients</td>
<td>No. of enrolled patients</td>
<td>Stop enrolment</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>Vinblastine + doxorubicin + irradiation</td>
<td>Hodgkin’s disease</td>
<td>NA versus April 2000 versus November 2001</td>
<td>Y FFS versus FFS</td>
<td>420 versus 348 patients</td>
<td>No. of enrolled patients</td>
<td>Stop enrolment</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>Gemcitabine + vinorelbine</td>
<td>Advanced non-small-cell lung cancer</td>
<td>June 1997 versus May 1999 versus July 2000</td>
<td>Y OS versus OS</td>
<td>240 versus 120 patients</td>
<td>No. of enrolled patients</td>
<td>Stop enrolment</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>Gemcitabine + vinorelbine + cisplatin</td>
<td>Advanced non-small-cell lung cancer</td>
<td>April 1997 versus April 1999 versus April 2000</td>
<td>N OS versus OS</td>
<td>240 versus 120 patients</td>
<td>No. of enrolled patients</td>
<td>Stop enrolment</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>Lipiodol iodine-131</td>
<td>Adjuvant in resectable hepatocellular carcinoma</td>
<td>April 1992 versus August 1997 versus March 1999</td>
<td>N Recurrence rate, recurrence sites, DFS, OS versus DFS</td>
<td>120 versus 43 patients</td>
<td>No. of enrolled patients</td>
<td>Stop enrolment</td>
<td>NT</td>
<td></td>
</tr>
<tr>
<td>Doxorubicin or ethoglucid</td>
<td>Adjuvant in superficial transitional cell bladder carcinoma</td>
<td>December 1979 versus December 1983 versus August 1997</td>
<td>N NA</td>
<td>NA versus 206 patients</td>
<td>No. of enrolled patients</td>
<td>Stop enrolment</td>
<td>NT</td>
<td></td>
</tr>
</tbody>
</table>

*aRegistration trial at the EMEA/Food and Drug Administration. A, active controlled group; ChlVPP/EVA, Chlorambucil, vinblastine, procarbazine, and prednisolone/etoposide, vincristine, and doxorubicin; DFS, disease-free survival; DSMC, Data and Safety Monitoring Committee; FFS, failure-free survival; FOLFOX, Fluourouracil, leucovorin, oxaliplatin; FU, fluorouracil; HER-2, human epidermal growth factor receptor; N, no; NA, not available; NT, control group receiving no treatment; OS, overall survival; P, control group receiving placebo; PFS, progression-free survival; RFS, relapse-free survival; RR, response rate; TtP, time to progression; Y, yes.
The mean study duration was 30 months (range 12–64 months). The median time lag between the end of enrolment (which coincides approximately with study termination) and the date of publication of the results in peer-reviewed journals was 22 months (range 3 months to 15 years).

## discussion and conclusions

Truncated RCTs reported as having been stopped early for benefit are becoming more frequent. Our findings highlight a consistent increase (>50%) in prematurely stopped trials in oncology during the last 3 years in comparison to whole period analysed (1997–2007).

Ethical reasons also play a role in the decision to stop a trial, since there is a responsibility to minimise the number of people given an unsafe, ineffective, or clearly inferior treatment. On the other hand, an interim analysis may also have drawbacks, since stopping trials early for apparent benefit will systematically overestimate treatment effects [32].

The studies analysed were formally well designed; all were randomised, controlled, on the basis of robust end points, and with a large sample size. Though criticism of the poor quality of oncological trials seems out of place, unfortunately early termination raises new important concerns. Our findings lead to a new awareness: oncological trials are now formally better designed than in the past, but they are too often stopped prematurely. This may cause harm resulting from unreliable findings prematurely translated into clinical practice. More than 78% of the RCTs published in the last 3 years with an interim analysis ending the trial were used for registration purposes. This suggests a commercial component in stopping trials prematurely.

Regarding the methodology used to conduct the interim analyses, sample sizes used to obtain the interim efficacy results varied widely. Substantial concern is raised by five studies which enrolled <40% of the sample planned for final analysis. It is obvious that the risk of overestimating treatment effects increases markedly when the sample is small. Therefore, it is very important to insist that a large number of events must occur before investigators or DSMCs examine interim data, although that cannot guarantee data reliability in any case. In addition, the heterogeneity in sample sizes indicates that these committees enjoy ample discretion in deciding whether to stop a clinical trial early for benefit.

Statistical simulations have shown that RCTs can overestimate the magnitude of the treatment effect depending on the timing of the decision to stop (i.e. the fraction of the total planned sample size or expected number of events) [33]. Furthermore, repeated interim analyses at short intervals raise concern about data reliability: this strategy risks looking as though it is seeking the statistical significance necessary to stop a trial. In addition, repeated analyses on the same data pool often lead to statistically significant results only by chance [34, 35].

If a trial is evaluating the long-term efficacy of a treatment of conditions such as cancer, short-term benefits, no matter how significant statistically, may not justify early stopping. Data on disease recurrence and progression, drug resistance, metastasis, or adverse events, all factors that weight heavily in the benefit/risk balance, could easily be missed. An early stop may reduce the likelihood of detecting a difference in overall survival (the only relevant end point in this setting) because of the small sample, the possibility of crossing-over the experimental drugs, and contamination with other treatments.

Interim analysis data should always be evaluated by a DSMC, which should be independent in the sense that the members should have no interests in the study and should not directly participate in it. Although the majority of RCT reports stated there was a DSMC, we believe that its independence should always be reported. Stopping a trial after an interim analysis is often motivated by ethical considerations. The large number of patients spared (~40%), as evidenced by our analysis, might support this. However, the relation between sparing patients and saving time and trial costs is also unquestionable and indicates that there is also a market-driven intent. Our findings show that only a very small percentage of trials (~4%) were stopped early for harm, i.e. serious adverse events, which is quite acceptable. Therefore, toxicity does not represent the main factor leading to early termination of trials.

Stopping a trial early does not guarantee that patients will receive the apparently beneficial treatment—assuming one
believes they should—if study findings are not immediately publicly disseminated. We found long delays between study termination and published reports (~2 years), possibly because of confidentiality concerns in light of the current regulatory process. If the trials had continued for these further 2 years, more efficacy and safety data could have been gathered. In addition, such delays further lengthen the time needed for translating trial findings into general practice.

The study suffers one main limitation: since there is no ‘standard’ for reporting interim analysis methodology in scientific journals, there may have been some heterogeneity in this respect and some information might have been missed, affecting the sensitivity of the analysis. This could be overcome if study protocols were publicly available and details of interim analysis were reported better in peer-reviewed journals, e.g. by adoption of the Consolidated Standards of Reporting Trials statement.

In conclusion, a decision whether to stop a clinical trial before its completion requires a complex of ethical, statistical, and practical considerations, indicating that results of RCTs stopped early for benefit should be viewed with criticism and need to be further confirmed. The main effect of such decisions is mainly to move forward to an earlier-than-ideal point along the drug approval path; this could jeopardise consumers’ health, leading to unsafe and ineffective drugs being marketed and prescribed. Even if well designed, truncated studies should not become routine. We believe that only untruncated trials can provide a full level of evidence which can be translated into clinical practice without further confirmative trials.

acknowledgements

We would like to thank J. D. Baggott for precious editorial assistance.

Disclosure of interest: Authors belong to regulatory and research institutions devoted to patients’ and national health services’ interests. Both are non-profit organisations. Authors did not receive any funds for conducting the present analysis.

references

32. Pocock SJ. When (not) to stop a clinical trial for benefit. JAMA 2005; 294: 2228–2230.