Debate about the benefits of chemotherapy as an adjunct to potentially curative surgery in patients with locally advanced gastric cancer has extended over the past several decades. There have been many negative studies, interspersed with a few positive ones; meta-analyses have suggested a small benefit of adjuvant therapy, but the widely varying regimens and populations of the studies included in these meta-analyses have limited their impact on this study design. Recently, however, two large randomized controlled studies in the United States and Europe have provided more persuasive evidence that adjunctive and/or neoadjuvant therapy can contribute to the elimination of micrometastatic and increased survival (1,2), but there remains substantial room for further improvement of outcomes.

In this issue of the Journal, Cascinu et al. (3) report on a study designed before the completion and reporting of the positive studies mentioned above. Their paper reprises a sadly familiar refrain in oncology drug development: a regimen shown to be promising in preliminary studies provides no improvement over a more “standard” and less intensive regimen when studied in a carefully designed and conducted randomized trial. The study of Cascinu et al. (3), conducted from 1998 to 2003, was designed to compare chemotherapy with an 8-week course of PELFw (weekly cisplatin 40 mg/m², leucovorin [LV] 250 mg/m², epirubicin 35 mg/m², 5-fluorouracil [5-FU] 500 mg/m², and glutathione 1500 mg/m²) with 6 months of 5-FU/LV (5-day course of LV 20 mg/m² and 5-FU 375 mg/m² every 28 days). The objective of the trial was to assess whether adjuvant therapy with the more intensive PELFw regimen would result in improved survival in patients with locally advanced gastric cancer. The investigators chose to use an active control arm based on a modestly positive meta-analysis (4) (although the authors of that meta-analysis were very cautious in their interpretation of their results and noted that adjuvant therapy should still be considered investigational). Final analysis, performed after the planned extent of follow-up and with a median follow-up time of 54 months, showed that both disease-free and overall survival were virtually identical in the two study arms.

What do we learn from this trial? First, we are reminded that many new regimens, despite initial enthusiasm based on preliminary studies, are ultimately shown to be ineffective or at least no more effective than other available treatments. Reasons for lack of effect could include inactive drugs or suboptimal regimens of active drugs. In this study, one might question the doses of chemotherapy administered. The protocol specified a cumulative dose over 6 months of 11250 mg/m² 5-FU, a venerable chemotherapeutic agent with modest to moderate activity in gastric cancer, for control patients, but only 4000 mg/m² 5-FU for patients undergoing the 8-week course of PELFw. The British MRC Adjuvant Gastric Cancer (MAGIC) trial using the same therapeutic components (epirubicin, cisplatin, 5-FU [ECF]) administered a much higher total dose of 5-FU, 25200 mg/m², together with doses of epirubicin and cisplatin comparable to those given in the PELFw arm of this study. (2) The substantially lower dose of 5-FU and shorter period of treatment may have reduced the potential benefits of this drug combination in comparison to the 5-FU-LV arm. (We cannot know from this design, of course, whether the PELFw regimen would have been superior to observation alone had such an arm been included.)
But a negative result could also be due to problems in the design or conduct of the study that obscured the true treatment effect. Many factors other than the adjuvant therapy administered may affect the observed outcome in patients with locally advanced gastric cancer, including (but not limited to) the extent of surgical resection, in particular the lymph node resection (D0 versus D1 versus D2); tumor location (gastroesophageal junction versus proximal versus distal); and pathologic characteristics (intestinal versus diffuse type). In addition, aspects of study conduct such as losses to follow-up, lack of adherence to assigned treatment, and suboptimal patient management may introduce excess variability and reduce the study’s ability to identify treatment differences.

The authors specifically addressed compliance with assigned treatment in this study. The tolerability of PELFw was low in this trial; although a similar proportion of patients on each arm completed the full course of treatment, only 9.4% (19 patients) completed the 8-week course of PELFw according to the planned dose and timing, compared with 33% for the longer duration 5-FU/LV regimen. The authors note that tolerability of postoperative therapy with similar regimens is generally low; the completion rate of postoperative adjuvant chemotherapy in the ECF arm of the MAGIC trial was 42% with higher cumulative doses of these drugs, but this figure does not account for those who required dose or timing modifications.

Another issue that can lead to negative results is selection of a patient population that is less sensitive to the gains that are potentially afforded by the new regimen. In this study, a large majority of patients underwent a D1 or D2 dissection, approaches that have become more standard over the past decade. These more extensive lymphadenectomies may decrease the rate of disease recurrence and lengthen survival and may well account for the excellent survival in both arms of this study. It is also possible that patients with more complete lymph node dissections derive less value from current approaches to adjuvant therapy. A previously published study (5) showed 5-year survival of 48% and 52% in locally advanced gastric cancer patients undergoing standardized D2 dissections without further treatment after surgery and patients receiving adjuvant therapy with the EAP regimen (etoposide, adriamycin, and cisplatin) followed by the Machover schedule (FU and folinic acid), respectively.

The observation of lower-than-anticipated failure rates in trials leading to midcourse sample size increases, extended follow-up duration, and, in many cases, inconclusive studies, is a long-appreciated phenomenon (6–8). Perhaps, the most extreme example is the Physician’s Health Study (9), in which the goal of evaluating the impact of daily aspirin therapy on fatal myocardial infarction had to be abandoned because the event rate after a number of years was only one-tenth of the expected rate. In this study, the low mortality unquestionably depleted the power of the study to identify any but very large improvements. The control arm was expected to yield a 20% 5-year survival in those treated with 5-FU-LV, and the study was designed to detect an improvement to 35% in the PELFw arm. To have the desired high power to demonstrate this difference, a total of 250 deaths needed to be observed during the 8-year course of the trial. With a 50% 5-year survival observed in both groups, only 163 deaths had been observed 8 years into the study among the 400 patients entered; many more years of follow-up would have been required to observe the number of deaths required for an adequate chance to demonstrate an improvement attributable to the new regimen.

Nevertheless, the data do not even hint at a trend to benefit that might have become statistically significant with a larger sample size and/or longer duration of follow-up; the data are more suggestive of equivalent efficacy [or inefficacy—without an observation-only arm one cannot be sure (10)]—of the two arms. The authors, while pointing out the reduced power, appropriately draw the straightforward conclusion that this intensive regimen is no better than the more “standard” regimen against which it was tested.

Results such as this should make us reflect on the typically unrecognized imprecision of what we think we know. As the modern era of clinical trials unfolded, many researchers were reluctant to randomize their patients between an unsatisfactory regimen then in use and a newly developed regimen that held out some hope of being better—maybe even much better. Some argued that randomized trials in this context were unethical and that patients should always be given the newer therapy, with results compared to those seen in earlier cohorts (11,12). This argument for the routine use of historical controls to evaluate new treatments ultimately was rejected by most researchers as failures of new, frequently more toxic, therapies, were successively documented and the limitations of the historical control method were fully recognized. Had Cascinu et al. (3) conducted a trial in which all subjects received the PELFw regimen, the observation of a survival rate vastly higher than had been anticipated could easily have led to the conclusion that this regimen was a huge advance over previous approaches to adjuvant therapy for gastric cancer; such a conclusion would likely today be leading many patients and their physicians to choose what is clearly a suboptimal and perhaps completely ineffective therapy instead of other regimens for which there is more definitive evidence of efficacy (1,2,13).

Why even raise this issue, given that most investigators now recognize the need to evaluate new regimens in well-designed and conducted randomized trials? Unfortunately, some legislators, pushed by patient advocacy groups, seem willing to ignore the lessons of the past and are pushing for an extensive rollback of regulatory policy such that drugs for cancer and other life-threatening diseases could be marketed without being shown to be beneficial in a rigorously designed and conducted trial such as this study’s authors have described (14). This trial provides an important reminder of how much patients have to lose should such efforts prevail.

References


