# *High-Dose Chemotherapy and Autologous Bone Marrow Support as Consolidation After Standard-Dose Adjuvant Therapy for High-Risk Primary Breast Cancer.*


## Peters, WP; Ross, M; Vredenburgh, JJ; et al.



## J Clin Oncology, 11 (6) : 1132-1143; June, 1993


**STUDY DESIGN AND CONDUCT**

**Study Aim**

The study aim was to investigate the use of high-dose cyclophosphamide, cisplatin and carmustine (HDCT) with autologous bone marrow support (ABMS) as consolidation after standard dose adjuvant chemotherapy (CAF) in the treatment of primary breast cancer involving 10 or more axillary nodes.

There is reasonable back-up for the trial objectives in the introduction, although it is interesting that radiotherapy was not considered as part of the protocol from the start.

There are several deficiencies in the stated aim. Firstly, there is no mention of the intended comparison group, which we later find are actually 2 arms from previous studies and 1 arm from a concurrent trial (CALGB 8541). Secondly, there is no statement of primary or secondary outcome variables to be measured or how these will be assessed. Thirdly there is no definitive statement indicating whether the protocol was prospective or retrospective (although it seems to be prospective). Fourthly there is no mention of the inclusion of radiotherapy or tamoxifen in the treatment schedule (although it is noted that radiotherapy was only included once the study was underway). These deficiencies should be rectified.

**Sample size and power**

There is no statement referring to sample size or power calculations, even though this is an important part of this type of study where there is comparison against historical controls (10). There is no statement of the expected improvement of the current protocol over previous studies. The authors do not state whether there was a pre-determined accrual time (eg to be ceased on a set date or after a set number of patients were entered). The authors need to include this information. Only 85 eligible patients were entered over 4 years.

**Patient selection**

**a. Source**

There is no statement defining the source of the patients. It is of interest that some patients were also entered onto another concurrent trial with "identical design and conduct". It would be interesting to know whether there were patients entered onto only one of the two trials during the period they were both running concurrently: if this occurred then this may point to referral bias or other unstated differences in protocols. The authors should clearly state the source of the subjects. For these patients, it appears that the present report is thus an interim analysis of an ongoing trial. The authors need to state whether this is the case, as well as describing the prospectively determined accrual times, follow-up times, and number of interim analyses to be performed as stated in this trial's protocol.

**b. Patient eligibility**

All patients with histologically proven stage II or III breast cancer with 10 or more lymph nodes involved at surgery were eligible. Patients were excluded if they had a history of prior cancer or treatment for the breast cancer.

There is no indication of whether there was independent verification of the histological diagnosis, and this ideally should be stated along with other measures of quality control. There is no age restriction for patients entered into the protocol, however the population is a young age group for this disease: it seems likely that there is an element of referral bias here. Also, a number of patients refused treatment, or were refused treatment by insurance companies: these patients may have differed in important factors from those patients included in the study (for example they may have been older or sicker). The 3 patients who were treated prior to referral and the patient who relapsed prior to commencing treatment may all have had aggressive disease: the poor prognostic patients may have been excluded. The authors should provide a comparison of these variables between the patients included in the study and those excluded. If the patients differ in important ways this will affect the ability to extrapolate results (extrapolation of results is already difficult due to the number of excluded patients: 17 of 102).

The authors do discuss the outcome of those patients who were not treated: 1 was lost to follow-up and 6 of the other 10 relapsed. No indication of actual therapies received is given.

**Treatment description**

Investigations performed are generally well documented, however there are several comments to be made. The investigations are extremely thorough, and as noted by the authors, are more likely to pick up metastatic disease than the historical studies.

Also baseline blood tests are not described and should be noted.

Several patients had delayed commencement of treatment which breached protocol: these patients were analysed with the other patients and were considered minor protocol violations.

Chemotherapy and support measures are well described. All patients received intended doses of chemotherapy. One area of concern is the "evolution" of the protocol with regards to support: 65 patients received colony stimulating factor primed peripheral blood progenitor cells. This alteration of the protocol occurred during the trial, and may have affected outcomes in an unknown way. It also leads us to question what other protocol changes occurred that have not been mentioned, and whether protocol changes were allowed for prospectively. In addition, there are several different schedules used for CSF administration: these do not appear to have been standardized in the protocol. The authors should provide the above information.

Another area of concern is the initial exclusion of radiotherapy as part of the treatment schedule and subsequent inclusion after high levels of loco-regional relapse. Of the 85 patients studied, 9 did not receive radiotherapy, and 3 of these relapsed, which is not surprising. Thus there is a subset of patients who had sub-optimal treatment, which does affect the results. The authors do note this in their discussion, however it limits extrapolation of results. Extrapolation is further hampered by the different radiotherapy doses used, as well as only some patients receiving radiation to the axilla: these differences may be very important considering the high risk of loco-regional failure in these patients. It seems that of those patients treated with radiotherapy, only 1 failed locally: this patient had "incomplete radiation therapy": does this mean that the axilla was not treated, or that treatment was aborted? The authors need to clearly state this, and discuss the difference in loco-regional failure between those that did and did not receive radiotherapy.

Another very important factor that has not been adequately considered is the use of tamoxifen for receptor positive patients. This does not appear to have been used in two of the control arms, and together with the radiotherapy and younger age-group may explain most of the difference in the results.

There is no mention of the surgical details, including type of mastectomy performed, level of nodes dissected, numbers of nodes excised, or the number of different surgeons performing the procedure. This information should be provided.

The treatment of relapses is not described and requires further discussion since these patients were included in results.

**Follow up**

Follow-up measures are fairly well described, however there is no mention of whether the same observer performed all follow-ups. The follow-up time is quite short for this group of patients, since relapse can occur many years after treatment. This is commented on in the discussion. The expected follow-up time of the trial before analysis should have been stated in the protocol, and requires a statement by the authors. The patients' last follow-up time appears to have varied. Patients also in the concurrent CALGB protocol seemed to have been assessed based on their last follow-up appointment (which could be up to 6 months previously), while those patients only on the current study were assessed on May 1, 1992 by "enquiry". The authors need to fully explain this term, and should preferably have assessed all patients over a similar time frame. There is no mention of whether any patients were lost to follow-up, and this should be stated.

**Outcome measures**

As previously discussed, primary and secondary outcomes have not been prospectively stated, nor with what they would be compared. The authors subsequently assess:

>actuarial probability of relapse at 30 months

>Kaplan-Meier estimate of survival at 2.5 years

>Kaplan-Meier estimate of event-free survival at 2.5 years

>Incidence of various toxicities

>Quality of life survey

>Median hospital charges (excluding some charges)

>Comparison of actuarial probability of relapse and event free survival with control arms, including comparison at 30 months with confidence intervals.

>Adjustment for certain prognostic factors.

The authors should state which of these endpoints were to be assessed prospectively. There are several problems with these endpoints which will be discussed later.

**Treatment completion**

Not all patients completed treatment planned:

4 patients were ineligible, 11 patients were not transplanted and 2 patients relapsed during CAF induction: these patients were

all excluded from the analysis, with no details regarding further management. As stated previously, more information is required regarding these patients.

It is not stated whether all patients that were planned to have radiotherapy or tamoxifen did have the full course of treatment: we are only told that 4 patients with lung toxicity required dose reduction or prolonged treatment time. This needs clarification.

**Side effects**

Treatment toxicity has been discussed at length in the text, however a table with graded toxicities is indicated for ease of comprehension. No attempt is made to list toxicity of control groups: this should be done to aid comparison between these and the current study.

**Comparison populations**

Three comparison populations were chosen: those from clinical trials of adjuvant chemotherapy in stage II breast cancer conducted by the CALGB during the past 17 years. These are not ideal populations to use as controls for several reasons:

>Many changes in cancer presentation, diagnosis and management (including support measures) will occur over this time period. Such changes are nearly impossible to quantify and will probably affect the comparison to an unknown degree. The authors acknowledge this.

>Inclusion criteria in the comparison studies was different from the present study, and thus certain subsets of patients were selected out for comparison: these subsets are still not exact matches in important variables, and this selection introduces further bias. There is no indication as to the source of the comparison subjects.

>The patients in the current trial have been extensively investigated compared to those in the previous studies: this is also acknowledged by the authors.

>The patient populations are different at least in terms of age, "more than 20 nodes" grouping, receptor status and local recurrences. Local recurrences are not defined. Two of the studies did not use radiotherapy, and for the other there are no clear details or numbers. The use of tamoxifen only occurred in the CALGB 8541 trial, and was not used at all in the other two. Other factors not considered include ECOG status, menopausal status, tumour size and internal mammary chain involvement which predict for outcomes measured.

>The adjuvant chemotherapy schedules received in the comparison studies were all different from each other, with unknown dose-intensities obtained.

>Of the most recent study (CALGB 8541), only 25 patients were compared to the present study, a point also made by the authors.

>There is no information regarding follow-up of the 3 studies.

The authors do comment on most of these problems in the discussion, and in fact state that "these uncontrolled data must also be interpreted with caution". These major concerns call into question how worthwhile the comparison is.

## ANALYSIS AND RESULTS

**Patient characteristics**

The current study's patient characteristics are displayed in Table 1, and those of the comparison groups are compared with the current study in Table 4. There are several important areas of concern:

>The numbers of patients in the current study listed with certain characteristics is different between the two tables. For example, Table 4 lists 33% of patients in the current study with 10-12 nodes, whereas Table 1 only states that there are 28 % in the same category. This requires clarification by the authors and correction, and calls into question the quality control measures instituted in the present study.

>There are some factors listed in Table 1 not compared with the other groups, including median number of involved nodes and surgical details. These should be included in the comparison.

>There is an obvious difference between the groups for several factors, most notably age. This has been commented on by the authors, but calls into question the possibility of referral bias: the authors need to comment on why their own study had such young patients.

>There are some important variables that have not been considered by the authors, including performance status, menopausal status, tumour size, level of axillary dissection and number of nodes excised: all of these may affect outcomes measured. These should be included in the comparison if possible.

**Statistical analysis**

The statistical methods used have mostly been referenced. The authors comment on the results of the comparison between the current study and the other studies being adjusted for the prognostic factors listed in Table 4. There is no mention of the

method used for this "adjustment", and these results are not provided. The authors should provide these results and reference the method of analysis used.

The actuarial probability of relapse at 30 months was 19% (95% confidence interval 0.10-0.44). The Kaplan-Meier estimate of survival at 2.5 years is 79% (95% confidence interval 0.64-0.88), and of event-free survival is 72% at 2.5 years (95% confidence intervals 0.56-0.82).

The analysis of relapse and survival is hampered by the short follow-up time: the median follow-up duration is only 2.5 years (30 months) and ranges from 16 months to 5.2 years. The short follow-up time needs further explanation by the authors: was the analysis performed at a prospectively determined time, or was it (as seems likely) determined during the study. If the latter is true then a degree of bias may have been introduced. There is a note at the end of the trial stating that an update was performed with a median follow-up of 3.3 years: actuarial event-free survival was 72% and overall survival was 77%.

The calculations of time intervals are measured from the first day of CAF: this is reasonable, however brings to our attention again the exclusion of patients from analysis (such as those two that relapsed during CAF). These patients should have been discussed more fully in the results.

The authors analyse event-free survival, a term that includes relapse or therapy-related mortality. As discussed in Devita (199, p.464), "it is generally unwise to exclude deaths from other causes": all deaths should be included. It is also noted that in these patients, measures such as overall survival are more important in terms of patient welfare. Overall survival is not compared between the current and comparison studies. These points should be addressed by the authors.

There are no actual figures given of patients alive with and without disease. Thus we are unable to discern whether all eligible patients were included in these results (nor if any patient was lost to follow-up). This information should be given, as should the number of patients who died from causes other than disease or toxicity, if any. It is stated only that eighteen patients died, 10 from toxicity and 8 of disease (as determined from the Table): we presume that no-one else died.

It is of note that five patients relapsed loco-regionally: 3 of these had no radiotherapy, and 1 relapsed in an unirradiated supraclavicular node. Thus the importance of radiotherapy in the patient group studied is huge. Of the seven patients who relapsed systemically, 5 appear to have died and 2 are classified as alive and disease-free after further treatment.

When comparing survival with the other three studies, only curves are provided of actuarial probability of relapse and event-free survival, with 95% confidence intervals determined at the median follow-up for the transplant study. It is noted that overall survival is not compared, and no logrank test is performed. Although the authors state that results were adjusted for listed prognostic variables, there is no indication of whether a regression model was used (as suggested by Gehan), and the data is not presented. Other important factors have not been adjusted for as previously discussed. The authors need to clearly state what analysis was performed and display the adjusted results, particularly since the confidence intervals overlap in the unadjusted comparison with the most recent trial.

Toxicity data is only listed as incidences, and only in the present study. As previously discussed, a table listing toxicities by grade should be given to enable easy reading. There are no toxicity data given specifically for surgery or radiotherapy. This is of importance, since there may be significant morbidity not reported, such as high incidences of lymphoedema. The comparison studies have not been assessed regarding toxicity, and no comparison has been made. Toxicity endpoints for the current study were discussed, and as part of a comparison with other studies, should be part of the comparison. The toxicity figures provided show a high degree of morbidity and mortality (including 10 toxic deaths out of a study of 85 patients). This is further reason to list toxicity by grade, and compare toxicities between studies.

Quality of life assessments are to be commended in this group of patients as an important endpoint to be assessed. Unfortunately the collection of data and subsequent analysis are not optimal. Data was only sought from 52 patients, with no indication how they were selected (except to say they were all more than 1 year post chemotherapy). All patients should have been assessed as part of the prospective trial protocol. Of these 52, only 43 patients completed the questionnaires, with no indication of the reasons for lack of information. All of these points should be clarified by the authors, and as a minimum there should be a comparison of the group that did have data collected versus those who did not. Preferably the authors should perform the data collection again, including all eligible patients on the study. There was no in-person assessment performed, and this may be a better way of data collection.

The comparison of the results of this questionnaire with another study of unknown patient make-up, treatment or follow-up is not adequate, and to say that the result was higher in the present study based on mean FLIC scores of 132 +/- 18 versus 126 for the old study is flawed. Either this comparison should not be made, or a proper comparison should be performed with the above information, and a sensible conclusion arrived at based on a statistical test.

In the analysis of costs (which again is to be commended), the authors exclude important charges which are directly related to the high-dose treatment (including bone marrow and PBPC harvesting, and physician costs). These costs add up to around $20,000 and do not include CSF costs. For the analysis to be meaningful, all directly related costs should be considered, not just a select few.

**Presentation**

In general, the presentation of data is adequate. The authors should provide the number of subjects at risk on the curves

displayed. The discrepancies in the figures between Tables 1 and 4 should be corrected, as well as including further information, as discussed. The authors should include a table of graded toxicities for the current and comparison studies.

## TRIAL CONCLUSIONS

In general, the authors have come to some sensible conclusions in the discussion, largely because they acknowledge many of the study's weaknesses.

The data show an apparent benefit for the entire high-dose consolidation program in terms of probability of relapse and event-free survival, however these results require confirmation in a prospective randomized trial before the therapy can be accepted for widespread use.

There is considerable toxicity associated with the high-dose treatment in terms of high morbidity and mortality levels. These levels may be reduced with advances in supportive care.

Quality of life appears acceptable compared with previously reported non-transplant regimens, and such measures should be used in randomized trials.

The selection of supportive care measures may have important effects on resource allocation, as measured by the hospital bill.

The toxicity, cost and complexity of treatment argue that the approach should only be used in major centres as part of randomized trials.

The American health insurance system is unfair.

Weaknesses acknowledged by the authors include:

>there was short follow-up

>the study was not randomized, historical controls were used

>morbidity and mortality were high

>there are problems with historical populations in general, including different patient selection, staging evaluation, age, hormone receptor status, dose-intensity, follow-up and unknown factors

>there were distinct problems with the subgroups chosen: small numbers in the concurrent CALGB 8541 trial, median age 10 years older, more patients with greater than 20 lymph nodes involved, no dose-intensity data for the 2 older trials, higher loco-regional failure in the comparison trials consistent with use of radiotherapy in the present study

>there were other factors that may have contributed to improved disease-free interval include loco-regional radiotherapy, tamoxifen and more thorough staging in the present study. These factors are not given as much attention in the discussion as they deserve.

The conclusions that the authors have arrived at are reasonable due to the weaknesses acknowledged, however the fact that these weaknesses were so obvious makes one wonder why the study was conducted as it was. Furthermore the authors could have equally come to the conclusion that when all of the weaknesses are taken into account, they could account for all of the differences between the current and comparison studies.

The statement about the cost of supportive care measures is even more poignant once the excluded charges are incorporated.

The statement about quality of life endpoints cannot really be backed up unless more information is provided as previously discussed.

The point that a randomized trial would be better than the present study design is well made on multiple occasions, and begs the question: why didn't they do one?

## RECOMMENDATIONS

This study does add insight into an area that has received a lot of attention in the journals. The study provides some very interesting information regarding protocols and toxicity for the high-dose regime used, as well as a warning to those who exclude radiotherapy when it is indicated. Although the study does have major flaws, most are actually discussed by the authors: I believe the trial does warrant publication. There are several areas however that I would suggest need to be reconsidered by the authors. These include:

>A clear statement of the outcome variables to be prospectively assessed, as well as their intended method of comparison. The prospective nature of the trial should be confirmed.

>The initial exclusion of radiotherapy from the protocol should be justified in the introduction, particularly since it was later included.

>Tamoxifen and radiotherapy should be mentioned in the trial abstract as forming part of the treatment regime. The importance of these two therapies needs to be emphasized.

>Determination of sample size, power calculations and termination of accrual should be indicated, and their prospective or retrospective nature stated.

>For those patients who are part of an ongoing trial, the prospectively determined follow-up times and number of interim analyses to be performed should be stated.

>Source of the patients should be stated.

>All quality control measures should be stated (eg histological confirmation).

>Details regarding reason for patient and insurance company refusal should be given, and a comparison of characteristics made between patients who were not eligible and those who were. A comment should be made regarding the possible exclusion of patients with aggressive disease (eg patients who were treated before assessment or those who relapsed before high-dose chemotherapy), older age or poorer health, and implications for extrapolation. The authors should specifically state whether there was any exclusion criteria for age, and if not, comment on the reasons that their study population is so young, including possible referral bias.

>State whether changes in the protocol (such as changes in support measures or use of radiotherapy) were allowed for prospectively. Discuss what effect on outcome the various support measures used might have, as well as explain why there was not a standardized regime for those measures used.

>Give more details, as previously stated, regarding radiotherapy and surgery: for example was there a standardized technique used, and which patients received axillary irradiation.

>State whether all patients who received radiotherapy and tamoxifen completed the intended treatments.

>Give details regarding the treatment of relapses.

>Give details as to the prospectively determined time of follow-up before assessment for analysis.

>Give details of what "enquiries" were made when gathering follow-up information, and consider re-evaluating all patients at the same time, in the same unit.

>State clearly whether any patients were lost to follow up.

>Re-evaluate the usefulness of the comparison with the historical controls (including the small concurrent study) in view of the many problems discussed. If a comparison is to remain, consider using more recent and similar historical controls if possible, and consider publishing the data once more adequate follow-up has occurred (ie consider delaying publication).

>If the authors wish to proceed with the current comparison groups, then provide a full list of all prognostic factors (as discussed), and present the results after they have been "adjusted" (and reference the regression method used). Present a comparison of overall survival data. Include death from all causes in the definition of event-free survival. Comment on the effect of excluding patients from the analysis on the results, or consider including them in the analysis. Indicate number of patients at risk on the survival curves. Clearly account for all patients who underwent treatment: at present we are left guessing. Comment on the confidence intervals of the adjusted analysis.

>Correct the discrepancies in Tables 1 and 4 (patient comparability information).

>Provide a table of graded toxicity data for the present and any comparison studies.

>State how patients were chosen to perform quality of life assessments, and describe why patients did not complete the questionnaire, and whether there were any important differences between those that completed them, and those that did not. Omit the comparison with the historical study unless more information is given (as discussed). Omit the statement that the score was higher in the present study unless there is a statistically significant difference between the two scores, and the two groups being compared turn out to be comparable.

>Include, or justify the exclusion of, the charges omitted from assessment.

>Discuss in more detail the importance of both radiotherapy and tamoxifen in the treatment regimen, particularly in the light of the high rate of loco-regional relapse when radiotherapy was initially not included.

>The authors should consider stating in their conclusions, that the multitude of potential biases and confounding variables in the present study may well account for all of the perceived differences between the present study and the comparison studies.