

# Comparison of Doxorubicin and Mitoxantrone in the Treatment of Elderly Patients with Advanced Diffuse Non-Hodgkin's Lymphoma Using CHOP Versus CNOP Chemotherapy.

Sonneveld, P; de Ridder, M; van der Lelie, H; et al.

*J Clin Oncology, 13 (10) : 2530-2539 Oct 1995*

## STUDY DESIGN AND CONDUCT

### Study aim

The authors state that there were two goals of the study. The first goal was to evaluate if full-dose CHOP is an effective and tolerable treatment in patients of age 60 or older with advanced intermediate and high grade Non-Hodgkin's Lymphoma (NHL). The second goal was to investigate if replacement of doxorubicin by mitoxantrone would significantly reduce the toxicity of the regimen without affecting its efficacy. In the introduction, the authors provide reasonable background to support the study's objectives. The authors propose to answer the questions posed by initiating a prospective, randomized multicentre phase III clinical trial of full-dose CHOP in patients aged 60yo or more with advanced intermediate and high grade NHL.

The main criticisms are in relation to the first goal. This objective is not alluded to in the paper's title, which seems unusual if this is the study's stated first goal. The authors fail to state at the outset what they will compare full-

It would seem that the two goals would be best studied by two separate trials. The first trial could either compare the effectiveness and tolerability of full-dose CHOP in patients older than 60yo with patients younger than 60yo, or alternatively compare the effectiveness and tolerability of full-dose CHOP versus low-dose CHOP in patients aged 60yo or older. The second trial would randomize patients to receive full-dose CHOP or CNOP, and compare efficacy and tolerability. An alternative way of studying both goals together could be by a three arm trial comparing a low-dose CHOP arm, a full-dose CHOP arm, and a CNOP arm.

Another criticism is that the authors do not state in advance the wish to determine prognostic factors or analyse differences in response rates, overall survival and relapse rates depending on these factors. In the results section of the paper these analyses have been performed, which leads us to question whether these objectives were asked retrospectively.

Thus the authors should state in the trial aim, the primary objective of the trial, as well as important secondary

### Sample size and Power

Power calculations prior to the study's commencement showed that 220 patients were required. However it is not clear what outcome the calculations were based on, nor the expected magnitude of difference in outcomes between the two groups. Sample size calculations are based on a single variable and are not valid for multiple outcomes (Altman,1991, p 454).

Instead of the expected accrual of 220 patients in 3 years, only 157 patients over 5 years were included (of which 9

The authors state that retrospective power calculations showed that the numbers in the trial were sufficient to demonstrate a difference in 5 year survival of 20% versus 40% between the groups at a P value of 0.05 with a power of 80%. However power calculations ideally should not be performed retrospectively, and this study does not quote 5 year survival data since it is not available (and the 3 year survival data show a difference of 26%

The authors should address the reasons that the expected sample size could not be attained, and recalculate the study power using the results actually obtained at 3 years (26 versus 42%).

### Patient selection

#### Source of patients

There is no reference to the source of the subjects entered into the trial. There may have been selection bias which could account for the poor accrual of patients. Source of the patients should be stated by the authors.

#### Patient eligibility

Eligibility criteria do not seem overly restrictive, except perhaps as discussed under organ function. Inclusion and exclusion criteria were as follows:

Consent : it is stated that all patients consented to participate, however it is uncertain whether consent was obtained prior to or after randomization. The refusal rate is not mentioned.

Age : age greater than 59 may be open to interpretation as including subjects in their sixtieth year (but not yet 60 years old). In the rest of the paper age criteria of 60yo is used.

Histological diagnosis : well defined and included second review. It is noted that there is missing histological information which requires explanation, and will be further discussed.

Stage : It does not appear that there were strict staging criteria required for entry into the study. Although all patients had staging CT scans only 94.6% of patients had a bone marrow biopsy: this inadequacy is not explained even though it is included as a "standard staging technique". Other investigatory procedures performed (in particular LDH) should have been stated in view of their subsequent analysis. The missing data for LDH is not alluded to by the authors and again one wonders whether this was analysed as an afterthought.

Bulky disease: defined as a biopsy-proven tumour mass of 10 cm. Although it is not stated we must assume that the measurements were calculated on CT scan. This should be made clear. As with several other characteristics, there is unexplained missing data relating to disease bulk: this requires explanation.

Performance status is well defined.

Organ function: Normal hepatic, kidney and pulmonary function are not defined. We cannot be sure if this was left open to individual physicians or laboratories to interpret, or whether there was part of a defined protocol. If strict criteria of normality were used then this may partly explain poor accrual, since a significant proportion of

Specific exclusion criteria are slightly ambiguous: we are unsure as to whether patients with any history of neoplasm are excluded, or whether only intercurrent neoplasms are excluded (and prior neoplasms may be included). It would seem unreasonable to exclude a patient, for example who had a cutaneous BCC excised 20 years ago with no recurrence.

It is interesting that despite these entry criteria, 9 enrolled patients were subsequently excluded from the analysis (including 2 patients who did not even have NHL!). This leads us to question the quality control features of the trial, another aspect that the authors should elucidate.

#### Patient stratification and randomization

Randomization was done by a "centralized randomization office" with no further details given. Randomization was unequal, with 76 patients in the CNOP arm and 72 patients in the CHOP arm. The mechanism of treatment allocation is not described and there is no information given as to the blindness of the trial, thus bias may have occurred. There is no comment regarding time from randomization to treatment commencement, however as there was one death that occurred after randomization and before the start of treatment this leads us to question whether the delay was

prolonged. There was no stratification for possible prognostic factors despite the later statistical analyses performed. Stratification avoids gross imbalances between the groups for known prognostic variables and helps prevent retrospective data dredging.

The authors should provide the reader with information as to the randomization procedures used and the degree of blindness in the trial.

### Treatment description

There are several areas that are not well defined in this section of the paper.

- A total of 30 patients were taken off study after 3 cycles because they were non-responders. They were treated with "other therapy"; there is no further elaboration. Since this group of patients are included in the survival analyses all treatments received are very important. By treating patients with unknown methods off study the trial is subject to bias and confounding factors. It also does not aid other clinicians to reproduce the study, or incorporate results into day to day practice. The authors should provide some information regarding further therapy for these subjects.
- The use of agents for control or prevention of side effects (eg. anti-emetics, colony stimulating factors) should have been discussed, particularly since side effects of treatment are analysed in detail in the results section and a judgement regarding tolerability of CHOP and CNOP is made. Further confounders have been introduced, and physician bias may have influenced the results.
- In the "treatment regimen" section of the paper we are introduced to the concept of "rapidly progressive disease". This concept is not defined in this paper even though it directly affects the treatment given, nor are we informed how it was assessed. This leads to two concerns. Firstly, it would be impossible for readers of the article to follow the correct treatment regimen without a definition of rapidly progressive disease. Secondly, if there was no written protocol to define this entity, bias may have been introduced by allowing each observer to interpret which subjects are to receive the first 3 cycles of chemotherapy in a shorter time. This more intense course of chemotherapy may influence efficacy and tolerability to an unknown extent. Results and side-effects of this treatment compared to the standard 28 day cycle treatment are not analysed. The number of patients with rapidly progressive disease in each treatment group are not listed, nor is this analysed as a possible prognostic variable. These concerns need to be adequately addressed by the authors.
- The dose of mitoxantrone chosen may be suboptimal, a point acknowledged by the authors in the discussion. If this is the case, some readers will argue that results with higher dose CNOP would have been superior to those obtained in the present trial.
- Procedures for dose reduction and delay are well described.
- More information could have been given regarding length of infusion times.

### Outcome measures

There are multiple outcomes measured and subsequently analysed, including: response rates (after 3 and 6 cycles), survival ( overall, 3 year and lymphoma-specific), disease free interval (median disease free interval, 2 year relapse free rates and 3 year disease free survival), dose intensity (completion of 6 cycles without dose reduction, mean normalized dose intensity after each cycle and cumulatively) and toxicity (incidence of graded toxicities after each cycle, incidence of toxic deaths, incidence of leucocytopenia/thrombocytopenia and median WBC/platelet nadirs following cycles 1,3 and 6 in 53 randomly chosen patients, incidence of congestive cardiac failure, incidence of reduction of LVEF of 15 % in 45 measured patients, and overall toxicity). Evaluation of prognostic factors was also performed for many outcome measures.

There are several comments to be made regarding the trial's outcome measures:

- It is desirable that one outcome measure be the main focus of the trial (and stated in advance). If there are one or two other important outcome measures these should be stated in advance, and be considered of secondary importance. Any interesting findings among the secondary outcome measures should be interpreted cautiously and perhaps be investigated in another study (Altman, 1991, p.454). If the authors believe that the most important clinical outcome to be measured is overall survival or disease-free survival then this should be stated. Secondary outcome measures that the authors may wish to analyse are toxicity or number of patients completing treatment. This trial does not state its main outcome measurement, and there are far too many outcomes measured and analysed, with the authors seemingly giving equal weight to all outcomes measured. One example is the evaluation and hypothesis testing of partial, complete and non response rates after 3 and 6 cycles of CHOP or CNOP, and overall best response. These seem to be given importance in the evaluation and discussion equal to the survival outcomes, despite the fact that response rates are less meaningful and more subjective.
- If the endpoints have been determined retrospectively then they are subject to bias. A good example is assessing 3 year survival: this should have been an endpoint stated at the outset of the trial (not determined retrospectively) in order to avoid bias. This also applies to effect of prognostic factors on outcome measures: they must be stated in advance. The authors must state in the trial aims/methods section the main and secondary outcomes that were determined prospectively in the trial protocol, and how they were to be assessed.
- There is no indication of the investigations required to document response rates. Also, response was documented after 3 and 6 cycles however some patients went on to receive 8 cycles. There is no mention of re-staging or assessment after 8 cycles.
- In assessment of survival and disease free intervals there is no indication of the regularity of follow up, nor follow-up investigations or other methods of assessment. There is no mention of why 16 patients were not assessable after 3 cycles: it is possible that there were inadequate staging tests performed on some of these patients. The authors should address these concerns.
- Toxicity outcomes have not been assessed in all patients. White cell and platelet levels have only been measured in 53 randomly chosen patients. It is not clear why these counts were not performed in all patients. Measurements of LVEF were only performed in 45 patients, with no indication of how these patients were selected. This should be better delineated. These small numbers are another important reason to use confidence intervals in the analysis.

### Treatment completion

As allowed for in the protocol, 30 of the 145 patients were taken off the study after 3 cycles of chemotherapy due to no response (21% of all patients). A total of 55 (of 71) CHOP and 46 (of 74) CNOP patients had 6 cycles of chemotherapy. Only 31% of the CNOP arm and 45% of the CHOP arm completed 6 cycles of planned treatment without dose reduction. There is no breakdown in Table 5 (which documents dose reduction) as to whether inability to complete planned treatment was due to patients being taken off trial, dying, suffering from toxicity or other reasons. Once again, there is no mention of the patients who were supposed to undergo 8 cycles of chemotherapy (63 patients in PR after 3 cycles), thus we do not know how many patients completed the planned treatment. These areas require addressing. Patients who dropped out due to death or those lost to follow up are described separately for each group in Table 3.

### Side effects

There is a detailed breakdown of graded toxicities using the common toxicity criteria for both treatment arms. The assessment of leucocytopenia/ WBC nadir, thrombocytopenia/ platelet nadir and LVEF was only performed in selected patients from each treatment arm. There is no

explanation as to why all patients were not assessed, nor how patients were selected for measurement of LVEF. The authors need to address this.

### Follow up

A total of 148 patients were randomized in this study. Only three patients out of 53 survivors were lost to follow up (median follow up for survivors was 23 months). The authors state that the median follow up time is only 14 months for all patients and 23 months for survivors only. Duration of follow up planned must be stated in advance (Devita, 1993, p.427). There is no indication of planned follow up for this trial, and this should also be discussed.

## **ANALYSIS AND RESULTS**

### Patient characteristics

Most baseline characteristics are well described for each group of the CHOP versus CNOP trial. Of concern however is the failure to describe the number of patients in each group with rapidly progressive disease. This presumably is an important prognostic factor, and certainly affects treatment given: this information should be provided. There is also missing data for histopathology, LDH level, performance status, bulky disease and extranodal sites. Reasons for missing data are not explained and it raises questions about whether these criteria were determined in advance: this should also be explained. In the patient characteristics listed, the CHOP and CNOP groups seem comparable.

With regards to the comparison of the CHOP arm and the retrospective study (Fisher et al) there is no attempt to list comparability of the two groups, therefore no firm conclusions can be drawn from this comparison. The authors should discuss the comparability of the two groups.

### Statistical analysis

Statistical analyses are described and referenced in the methods section of the trial. Power calculation methods are not adequately described or referenced. This is an area that should have been better discussed in view of the sample size being inadequate after initial calculations but apparently found to be acceptable with retrospective calculations.

The tests utilized are generally appropriate for the results to be analysed, however the use of confidence intervals instead of (or in addition to) hypothesis testing are more desirable (Altman, 1991, p.175). Confidence intervals can show whether a difference is actually clinically relevant. Also, the use of P values alone can hide the fact that the study was too small, but a wide confidence interval can indicate a lack of useful data. This may be particularly relevant in this study where there is a question regarding adequate patient numbers. The authors should present their results as confidence intervals with or without P values.

When comparing the CHOP and CNOP arms of the present trial, far too many outcome variables have been hypothesis tested. As previously stated, one or two main outcomes should be analysed using confidence intervals with or without hypothesis testing. The overuse of hypothesis testing leads to a high likelihood of generating significant results which obscure the important outcomes analysed. The authors should state the prospectively determined outcome variables and analyse these. Other results should be listed without excessive hypothesis testing.

When comparing relapse rates, there are 16 patients not assessable, yet only 3 are mentioned in the text. This needs correction, and explanation as to why the intended protocol was not adhered to. Quality control procedures are again in question.

Comparison of survival data using survival curves and deriving median survivals from these curves is appropriate. The use of hazard ratios is also appropriate and complements other survival function analysis (14). Hazard ratios are not utilised in the present study, and confidence intervals should be utilized. The use of 3 year survival should be avoided (Altman p 386), especially if this was not stated in advance: this type of analysis wastes information and may introduce bias. In addition,

although 3 year survival is quoted, the median follow-up time is only 14 months (and 23 months for survivors): the authors should consider quoting survival figures for the minimum follow-up time. Interestingly, at 3 years only 17% of CHOP and 13% of CNOP patients were alive and disease-free (no significant difference), a fact not further discussed by the authors.

The use of the intention to treat principle is appropriate, however as previously discussed there is no mention of the exact treatments or outcomes for the 30 patients that were taken off trial after 3 cycles of chemotherapy.

When comparing the current trial's CHOP arm with another study's CHOP arm in the discussion, the authors have quoted CR and OS rates from each study as percentages and stated that these are comparable. There is no attempt to compare the study sizes, prognostic factors, treatments or to perform any statistical analyses. If they wish to proceed with such a comparison, the authors should attempt to determine comparability between the studies.

With regards to analysis of prognostic factors, there are several areas that require attention:

- It is generally desirable to incorporate major prognostic factors as stratification variables in the randomization (Devita, 1993, p.434), and subset analyses should be specified in advance as defined secondary analyses. The authors should describe those analyses that were prospectively determined, and refrain from those done post-hoc.
- The authors have performed multiple subset analyses which increase the probability of finding a statistically significant difference by chance alone.
- It is generally not valid to subset the analysis by characteristics measured after the start of treatment (Devita, 1993, p.435): the authors tested the effect of time needed to achieve CR as well as normalised dose-intensity on survival. This analysis should not be performed.
- Once again, the missing prognostic factor data (eg LDH levels) is noted, and requires explanation.

### Presentation

Statistical material is generally well presented in the text and as tables and graphs. The article does seem a little cluttered by tables and graphs (reflecting the large number of outcome variables assessed). Some of these (for example Table 7, displaying toxicities) are quite complicated. Altman (1991, p.386) suggests curtailing survival curves when there are only 5 subjects still at risk (the authors display theirs to 60 months when there are only 4 patients at risk).

## **TRIAL CONCLUSIONS**

The authors draw many conclusions from the multiple analyses performed. These include:

a. The response rate in elderly patients is comparable to that achieved in younger patients and complete responders have a good probability of long survival. However survival in this patient group is observed exclusively in patients who respond to therapy. These conclusions are based on the following observations:

- Using a 4 week CHOP schedule, the CR rate in elderly patients with aggressive NHL is 49% which is comparable to an unselected group of adult patients treated with 3 week CHOP (CR rate 44%).
- Overall survival rate at 3 years was 41% compared with 54% in patients of all ages.
- Of those patients who achieve CR, the prospect of remaining in remission after CHOP is 54%.

b. Doxorubicin and mitoxantrone were equally tolerated by elderly patients with aggressive NHL. This was based on:

- The incidence and grading of toxicity by CHOP and CNOP are comparable, except for nausea ( $p=.02$ ) and alopecia ( $p$  less than  $.001$ ), which were less severe with CNOP.
- Dose intensities could be maintained to the same level in CHOP and CNOP patients (31% of CNOP patients v 45% of CHOP patients completed 6 cycles without dose reduction,  $p=.09$ ).

c. CHOP has greater efficacy than CNOP in the treatment of elderly patients with aggressive NHL.

- CHOP achieves more responses than CNOP (PRs after 3 cycles 62% v 35%,  $p=.002$ ; for overall best response, a CR was found in 49% v 31%,  $p=.03$ ).
- Overall survival of CHOP patients (median 26 months) was significantly better than CNOP patients (median 12 months):  $p=.029$ . 3 year survival was 42% v 26%. Lymphoma specific survival was significantly better with CHOP ( $p=.034$ ).
- Elderly patients who have a slow response to CHOP should be treated with additional cycles, and the better efficacy of CHOP in elderly patients appears to be associated with completion of the scheduled therapy. This is based on the observation (but no statistical assessment) that many additional patients treated with CHOP who achieved a PR after 3 cycles ultimately had a CR, while few patients in PR after the third cycle of CNOP attained a CR.
- Further dose escalation of chemotherapy in elderly patients appears possible. This is based on the observation that adherence to the chemotherapy schedule was comparable between patients of different ages, and that toxicity could be reduced using supportive methods such as G-CSF.
- The final conclusion is that CHOP is recommended for the treatment of high-risk NHL in elderly patients.
- There are several important problems with the trial as it is presented. These problems make it difficult to draw firm conclusions from the results. Some of the main areas of concern include:
  - The study is really only designed to answer the second stated objective. In doing so it will provide some interesting information relating to the first objective but a second trial would be required to answer the question posed. Comparison of the high-dose CHOP arm with a previously reported low dose CHOP study without any attempt to determine subject comparability, treatment details or toxicity of the old study, or to statistically analyse differences is suboptimal.
  - There is a question regarding adequacy of patient numbers. Power calculations determined at the outset were ignored and then recalculated retrospectively in an attempt to justify the sample size obtained. The authors believe that the size obtained (which is about 50% of that initially calculated) is adequate. Power calculations are performed for a single endpoint to be analysed: this endpoint is not delineated by the authors, and multiple endpoints were analysed. The authors appear to have used a one sided test when a two sided significance level of 0.05 is widely accepted as standard (Devita 1993, p.428 ). Using a two sided test (as described in Devita), patient numbers do not seem adequate to detect a difference of 40% v 20% 5 year survival as claimed by the authors. In fact, the difference in survival is actually 42% v 26% at 3 years: this would require even greater patient numbers.
  - The patients were enrolled over a five year period which is also barely acceptable, and there are unanswered questions about why accrual was so poor. The source of the subjects is not described, thus selection bias may have occurred.

- There was no stated main outcome measure, and indeed there are far too many outcomes measured, overshadowing the most important ones. Outcome measures such as response rates are less meaningful and more subject to bias than survival.
- The excessive use of hypothesis testing increases the likelihood of finding significant results by chance alone. The use of confidence intervals is far preferred in any case as previously discussed.
- The dose of mitoxantrone used may be suboptimal, a point made by the authors in the discussion. If this is true, it may be argued that this could explain why the response rates and survival of the CNOP arm are inferior to the CHOP arm.
- In the assessment of response rates, toxicity and ability to tolerate planned treatment there is no mention of the 63 patients who were supposed to have undergone 8 cycles of chemotherapy as stated in the trial protocol. Assessments only seem to have been carried out up to the sixth cycle.
- Treatments were not well described. Patients taken off study had "other therapy" which was not further discussed. This is important as these patients were included in survival results. Rapidly progressive disease and treatment of side-effects/ supportive agents are not defined.
- Patients were included with inadequate staging and assessment of prognostic factors, leading us to question how tight the protocol was.
- There is no detailed account of randomization or degree of blindness, and thus bias may have occurred.
- Follow up is short (median 14 months, 23 months for survivors only)
- The analysis of prognostic factors is probably retrospective with no stratification and missing data. There are multiple subset analyses. Rapidly progressive disease is not recognised as a prognostic factor despite the different treatments received by such patients: this adds another confounding variable.

## **RECOMMENDATIONS**

This trial examines an interesting aspect in the management of elderly patients with advanced high-grade NHL, however does suffer from several flaws as discussed. I believe the study should be published, but I would suggest the authors review several aspects before its publication:

- The trial's prospectively determined objectives and primary and secondary endpoints to be studied need to be clearly stated in the aim (including assessment of prognostic factors and anticipated subset analyses). Their method of assessment should be stated in the methods section.
- An explanation is required regarding the very poor accrual of the trial, and patient source should be described. The retrospective power calculations should be recalculated by incorporating the actual results obtained, ensuring a two sided test is used.
- The source of the patient should be described.
- Trial randomization procedures and degree of blindness need to be clearly stated.
- The median follow-up time of the trial is short and needs explanation. The authors should state the prospectively determined follow-up time for the trial, as well as better describing follow-up frequency and investigations.
- The authors should explain why patients have been included without adequate staging and information regarding prognostic variables.



- Quality control measures instituted by the authors should be discussed.
- Authors should comment on the possible exclusion of patients with metastatic disease in the liver.
- Treatments (including support measures) should be better described and include discussion of management of patients taken off trial.
- The authors should restrict their excessive analyses to the stated outcomes to be measured. All analyses, where appropriate, should include 95% confidence intervals. Conclusions can be drawn from the re-analysis. When discussing 3 year survival figures, the authors should comment on the short follow-up: alternatively, survival could be analysed using minimum follow-up time. Hazard ratios could also be determined. In addition, when discussing the two treatment arms, a comment should be made regarding the finding that at 3 years only 17% of CHOP and 13% of CNOP patients were alive and disease-free (with no significant difference). The analysis of the effect of CR and NDI on survival should be omitted. The analysis of the effect of other variables on survival should be restricted to those that were prospectively intended. The authors should state how patients were selected for measurements of LVEF.
- In the assessment of response rates, toxicity and ability to tolerate planned treatment there is no mention of the 63 patients who were supposed to have undergone 8 cycles of chemotherapy as stated in the trial protocol. These patients require assessment.
- The text should be corrected to state that 16 patients were not assessable for response with an explanation of why this has occurred.
- Patients who were taken off study should have their "other therapy" discussed. This is important as these patients were included in survival results. Rapidly progressive disease and treatment of side-effects/ supportive agents require definition. More information is needed with regards to number and distribution of patients with rapidly progressive disease, as well as the possible prognostic significance of this entity.
- In the discussion if the authors wish to proceed with the comparison of the high-dose CHOP arm with a previously reported low dose CHOP study, they should attempt to determine subject comparability, treatment details and toxicity of the old study. This would at least provide the reader with some information so as to make up their own mind regarding the assessment of comparability of the two treatments.
- The authors should consider resubmitting the trial after longer follow-up data is available, particularly to determine if there is any prolonged survival advantage.