Randomised trial of palliative two-fraction versus more intensive 13-fraction radiotherapy for patients with inoperable non-small cell lung cancer and good performance status.

MRC Lung Cancer Working Party - Bolger J.J, Hopwood P, Blehan N.M, Cartmell J, Girling D.J, Machin D, Stephens R.J and Bailey A.J.

Clinical Oncology 1996, 8:167-175.

BRIEF SUMMARY:

This is a randomised trial with two treatment arms:17Gy in 2 fractions (F2) and 39Gy in 13 fractions (F13). The endpoints are survival and quality of life. The study found that survival was better in the F13 but that palliation was more rapid in the F2 group.

AIM:

The aim of the study is to determine whether there is any additional benefit for inoperable non-small cell lung cancer patients by increasing the dose above the minimum required to achieve palliation of symptoms. The endpoints are not clearly stated but they have analysed both survival and palliation of symptoms. This study is restricted to those with good performance status, as defined by ECOG 0-2. These endpoints are appropriate measures of the question asked.

TRIAL DESIGN:

Patient accrual:
It is not stated how the patients were accrued. The population the study group are taken from is unresectable patients with non-small cell and non-carcinoid tumours of the lung with good performance status:ECOG 0-2. Patients had locally advanced but not metastatic disease. It is reasonable to exclude small cell and carcinoid histologies as these have quite different prognoses than non-small cell lung cancer. While the rationale to include those patients with only good ECOG status is reasonable, this excludes the majority of patients with lung cancer because this is a group with a high proportion of co-morbidities. This wastes a high proportion of available patients. Those with metastatic disease could have been used to boost numbers in the palliation of symptoms arms, and those of good performance status could have been used to look at survival.

Diagnostic tests used to define metastatic disease were not stated and this is an omission. Patients with prior malignancies were excluded except for those with basal cell carcinoma and carcinoma in situ of the cervix. While these criteria are almost standard, the life expectancy in these patients is so poor that widening of inclusion criteria eg to include patients with previous malignancies inactive for 5 years would probably not have biased the results.

It is not stated what treatment those excluded from the study received ie what is regarded as the standard arm in this group, although the introduction refers to previous MRC trials (1, 2) as standard practise.

Patients were randomised by telephone for consultant, performance status, histological type and sex. These are reasonable stratification categories. It is important to stratify for consultant to minimise differences in placement of field borders if there will be no central review or quality control of treatment planning as in this case. Patients are not randomised for weight loss. This is an important prognostic factor (3) and not to randomise for this is an omission.

Treatment Description:
Definition of treatments are well documented with energy, dose prescription point, air correction and absence of spinal cord shielding stipulated. Field configuration was described as opposed fields (not necessarily AP/PA) with dose prescription to midplane. The target is stated to be the primary and mediastinal lymph nodes. It would be more helpful if more information regarding the field configuration was given, allowing the reader to gauge spinal cord doses achieved as this is a major toxicity of concern. For the higher dose group a reduction in the dose to 36Gy is allowed at the discretion of the attending physician It is sensible to put this clause in, as the doses are above spinal cord tolerance when converted to biological equivalent dose and some physicians would be reluctant to treat patients with this regimen. Reducing dose allows for individual physician tolerances for spinal cord dose, thus greater accrual is

achieved. The equivalent doses are 62.4 Gy for late effects of 39Gy/13# and 57.6Gy for late effects for 36Gy/12#. Guidance for an upper limit for field volume of 200cm^2 is stated.

It is not stated if patients received any adjunctive treatments, except that steroids were recommended if there was evidence of superior vena cava obstruction.

The dose in the treatment groups is not regarded as standard in much of the world, but reflects UK patterns of practise and as such it is valid to include 17Gy/2# as a palliative standard. It is valid to use this as a baseline fractionation. Twenty Gray in five fractions or thirty Gray in ten fractions are probably more commonly used elsewhere in the world. It is prudent to design a study in which the two groups are as different as possible (4).

Followup:
Because the study measured quality of life issues as well as survival, follow-up needs to be more frequent than in a study assessing pure survival. Followup was done at 1, 2, 4, 6, 9, and 12 months with 6 monthly followup thereafter,with treatment related morbidity assessed by asking the patients to use a score card each evening. Patient status was not verified with professional contact during the first month after treatment.

Later followup was by clinician assessment and documents the treatment details, status of the primary and any evidence of any metastatic disease, but not morbidity. In addition patients are asked to perform a modified Rotterdam Symptom Checklist (RSCL) (5) - modified by adding questions specifically related to lung cancer symptoms and the Hospital Anxiety and Depression score (6). This score has been developed for use in general hospital outpatients, and was verified by psychiatric interview. The use of such a scale is to be commended.

All tools including psychometric scales and quality of life scores need to be tested that they accurately represent the status for each patient. They need to be valid across cultures, populations demographics, and diseases(7). Any additions and modifications to formally assessed tests may not also be accurate and cannot be verified as such without a formal test to assess this. The authors reference a thesis on the application of the RSCL in lung cancer, and state this as the justification for adding the extra factor of restlessness into the analysis. The results of this thesis are not expanded upon. The authors do not state the cut-off points used in the analysis of the HAD scales. The original paper found that the level of false negatives and positives changed with the scoring thresholds used. The RSCL was also used to measure psychological distress. It would be helpful if a comment was made on the correlation between the two methods.

For those not familiar with the RSCL and the HAD it would have been more helpful if the authors gave a brief summary of the scales used and how the scoring system worked.

No definitions of clinical relapse have been given and there are no stipulation of the minimum requirements for the diagnosis of relapse, either distant or local. This becomes important when the results are reported.

Statistical Design:
Power of the study was stated to be 5% significance for a 15% increase in survival with a 90% power. The authors calculated this would detect a 20% increase in the main side effect of radiotherapy which is dysphagia. There is no allowance for reduced compliance, dropouts or withdrawals in the sample size calculation.

It is extremely optimistic that the regimens would result in a 15% increase in survival and it may have been more helpful to have also calculated the sample size based on a lower increase in survival and a lower power. In addition, it would be desirable to detect a smaller increase in the occurrence of dysphagia, as we are dealing with a predominantly palliative regimen where the main issues are a trade-off between side-effects, and an increase in the time to progression.

The study actually accrued earlier than anticipated and therefore underwent in interim analysis which suggested that the treatment effect would be less than expected. Patient accrual was increased to 500 to allow for this effect. This was a responsible solution. The authors should have included the new power of the study. If the trial was stopped when expected, then the null result would have not been clinically

useful. In performing an interim analysis, the authors have assumed the proportional hazards assumption. This is not presented as verified.(8).

Kaplan-Meier survival curves and log-rank tests are appropriate for the analysis of survival curves. Presentation of single hazard ratios and confidence intervals are good practise, but again, the presentation of single hazard ratios is only valid if the proportional hazards assumption holds. Inclusions in the final analysis are from intention to treat until censoring or death. This study, because it looks at survival as a first endpoint, may have been better served by including only patients who completed treatment. All patients randomised for treatment should be included when assessing treatment morbidity as this also includes patients who fail to complete treatment due to treatment morbidity.

Quality Control:
The histology is double checked blind by an independent histopathologist. No comment is made of what diagnosis was made or how a consensus was reached if the reference pathologist disagreed with the referring pathologist.

There is no reference to quality control or assessment of the simulator films by an independent party. No double checking of the staging information was undertaken by the randomising centre. There is no stated quality control of the data entry. Collection of results was by way or reports at defined intervals after treatment, and included factors related to disease progression. It is not stated how the patient generated report cards were collected and the frequency. It is not stated whether there was any verification of the patient generated report cards and HAD scales by patient interview to check patient comprehension of the cards or validity of the data. Compliance is reported for both patient and clinician reports but it is not stated how the missing data is handled in the analysis of these reports. The report states that the COMPACT program managed the trial data. It is not stated whether this was the statistical package used or whether this was merely the data management package.

The lack of quality control in this study is a serious flaw.

Ethics committee approval, and individual patient consent was sought, and this is appropriate.

RESULTS ANALYSIS:

All numbers randomised are accounted for. The histology was confirmed in 454 patients. No slides were available in 41 patients and impossible to assess in 14 patients. Therefore 55 patients in the study have no confirmed diagnosis. In addition, 15 patients had clear criteria for exclusion, based on carcinoid, small cell, or in situ disease but were included in the analysis. All of these patients were evenly distributed across the two treatment groups. This does not make the inclusion of these patients valid and the last 15 patients at least should have been excluded. There could also be made a case for exclusion of the 55 patients with no confirmed histology. Practically there will be cases of undifferentiated carcinoma which will not fall into any of the previous groups. This does however allow a source of bias. I am concerned that the rate of impossible histology totalled 10% of the population. In addition to the possibility of introducing patients with histology other than NSCLC primary, there is the possibility that cases of non-lung cancer were included. There is the potential to have a significant impact on the results both of the survival and palliation endpoints.

The groups were equal in terms of sex, age, histology, performance status and presence of superior vena cava obstruction. At the bottom of table 1 of a scale of "overall condition" is included. This is subdivided as excellent/ good/ fair/ poor/ very poor and not known. There is no reference as to where this data was derived. It is superfluous information as the ECOG status is also presented in Table 1.

Treatment Delivery:
The authors report that 97% of F2 group and 89% of F13 group received treatment according to the protocol but that these numbers included patients who had a "minor deviation" from the protocol. Minor and major deviations are not defined in the methods section. There is no reference as to who or how treatment deviations were defined. Neither is description of the treatment deviations given in sufficient detail to allow the reader to make a decision. . For example, the F2 group included 7 patients who experienced a 2 week delay in between the 2 fractions, or only had half the treatment or received no treatment at all. Similarly the F13 group 27 patients received either prolonged treatment, altered fractionation or no treatment. All of these patients were included in the final analysis. In the first instance it would be sensible to remove those patients who received no treatment from the analysis as they will bias the survival. It would be helpful to the reader to give more information on the variation in doses actually received in each group and the fractionation so that the reader could decide if the inclusion of these patients was justified. It would also be helpful to give information of the size of the fields planned (+/- CI) as this will have an impact on the morbidity of the treatments. The authors should state why the

treatments were delayed, whether because of co-morbidity, disease complications and morbidity or because of machine availability. One alternative to this is to present the results with and without these patients added to see if the treatment deviations made any difference to the results.

A small number of patients also received radiotherapy to other sites of disease, or chemotherapy or endobronchial laser treatment. While these patients will occur in our practises and I do not think they should be excluded from the study, the authors do not state how they handled this with respect to measuring the QOL measures ie a patient who also has RT to the hip for bony metastasis could have improved mobility due to reduced pain in the hip or because of reduced shortness of breath due to the thoracic radiotherapy. They state a small number of patients also required additional radiotherapy to the chest, more in the F2 group. The authors do not state whether this was for in-field relapse, and if the difference was statistically significant ie. whether the further radiotherapy was given for failure of primary radiotherapy to control the disease, or whether it was given for geographic miss, or metastatic disease.

Kaplan-Meier survival curves are appropriately presented, although censored observations are not clear, this may be because of the scale of the graph. The median followup and range for the patients is also presented. The log-rank test demonstrates a difference between the two groups with a hazard ratio (HR) of 0.82 (95% CI 0.69-0.99) with a p value of 0.03. This means that the chance of this result occurring purely by chance is 0.03. Probably of more interest to the patient, the median survivals were 2 months longer in the F13 group. Medians are a reliable estimate of the middle measure of survival if the distribution follows a normal distribution, in which case the arithmetic mean is usually as good a measure. If the population is skewed, as most survival curves are, then this misleading. The geometric mean or the mode is better. The authors do not present enough information about the survival rates for the reader to assess the validity of the use of the median in this situation(9). The use of 1 and 2 year survival rates given by the authors is more informative and does not presume any information about the distribution of the survivals over time. The authors go on to state that "the hazard ratio was unaffected by stratified analyses for the stratification variables "(viz histology, consultant, performance status, sex) but contradict themselves when they say that performance status and histology had prognostic influence on survival. They also say that subgroup analysis by performance status suggested a trend towards better survival in the F13 group with HR = 0.68, 0.81, 0.91 for ECOG status 0, 1, and 2 but the log-rank test was not significant. They do not elaborate on the magnitude of the effect found for histology. They state that large cell and "other" histologies showed a poorer survival. Previously I mentioned the large numbers of patients in this group, and the intentions of the authors to exclude this group from the study and their subsequent failure to do so.

Treatment Failures:
The authors give their information as local and distant recurrences which is valuable information as regards the mechanism of failure. The results are reported in tabular form (Tables 2 and 3) and also in the text of the paper. This is unnecessary repetition as the table is quite clear. Because no criteria for relapse was defined, I am unsure of the significance of the 2 categories for local relapse: viz. Definite relapse vs Definite or suspected relapse. The addition of the suspected cases made little difference to the HR or the confidence intervals however. The results are presented as 6 , 12 and 24 month recurrence rates which are more informative to the clinician. It is not clear at what time point the hazard ratios are calculated. The reporting of site of first failure is also ambiguous. The numbers of patients reported as failing in the chest first for the F13 group exceeds the number presented as local failures in Table 2, implying that a proportion of these patients failed after 2 years. Total numbers of failure s are not given, and this would have helped to confirm my conclusion. Likewise, for distant metastases. The numbers of patients who failed with distant metastases first in the F2 group exceeds the 24 month distant relapse rate. No numbers of total failures are given. It is not stated for the small number of patients who relapsed at both sites at the same time whether they were included in both tables of local and distant relapse. It is useful to see the pattern of distant relapse subdivided for the 3 major organs. While this is not information expressly sought in the initial design of the trial, it is information which has been collected and which may be useful. The natural history of diseases have been known to change over time. It is a useful control in this case for the large number of unknown histology patients. If there was contamination of the patient population with small cell lung cancer patients we would expect a higher number of metastases with a much shorter survival. In this case the number of metastases are within the limits expected for non-small cell lung cancer.

Palliation of Symptoms:.
Palliation is defined clearly in the text as disappearance of symptoms or improvement of symptoms by 1 or more points on the RSCL score.

Compliance of the RSCL and clinician forms was assessed and this is good practise. Patient compliance was 65% and clinician compliance was 70% This is a high rate of non-compliance, but probably of the order expected if there is no provision for an on site coordinator of data collection who had facilities available to chase up missing reports. Although it is not stated it could be that the patient scales were posted directly by the patient to the coordinating centre. The report goes on to state that of those clinicians reports received, 89% of the RSCL scores were present. This is superfluous information as no attempt at correlation of the patient and clinician scores. The authors do not state if all of the data, incomplete and otherwise was presented in the results. Most importantly, the authors state that there was no difference in the compliance rates of the two groups.

The RSCL scores were analysed using Kaplan-Meier estimates and a log-rank analysis presented which is the appropriate test in this instance.

The presentation of crude scores are influenced by the survival effect. Because the averaged results for the whole group is presented in Figure 4 and the symptoms are closely related to the disease there is possible confounding due to intercurrent illness, and related neoplastic events. These should be addressed by the randomisation. The reader must beware that to interpret any difference as solely due to the treatment could be wrong.

Patient assessment of activities of daily living (ADL) are presented as median scores and interquartile range. What is obvious from the graph is that the populations have a very large range, and that the populations are skewed, as would be expected. The changes with time of the ADL assessments are presented in graphic form. Although the authors point the reader to lower scores in the F13 group with time, there is no statistical test of the difference between the two groups. It would be more correct to say that the results show a trend towards a lower ADL score in the F13 group, but that the results do not approach significance.

Assessment of psychological distress are presented as a graph of patients with a RSCL score of 12 or more, with percentages of patients printed above each bar on the graph. Again, the authors state that there was a difference on the two groups but do not provide any statistical verification for the statement. It may be a reflection that as some patients moved out of the RSCL>12 group, other patients moved into it as they coped with different crises in their disease. It would perhaps be more informative to look at change in RSCL score for each patient rather than the group as a whole.

Diary card assessments were presented as graphs for each symptom of percentage of patients with each symptom against time after treatment. The main concern with this is that given the low compliance rate (69%) it would be better to present this data as percentage of responses received . It is not clear if the was done. As the absolute numbers of patients are not given, there is no way of verifying this from the paper. There is less concern over the difference between the two groups as equal numbers of responses were received in each group. No statistical measure of the difference is given.

Assessment of adverse effects by clinicians only was reported in the text. It is unsure how they were processed and the implication is with the presentation of hazard ratios and confidence intervals that the results have been analysed by the Kaplan-Meier method, which is not appropriate.

There is also included in the results a comment regarding radiation myelopathy. It is an omission that the authors do not intend in their treatment design to monitor this parameter as the dose in the F13 group approaches a significant risk for myelopathy

and this has serious consequences. What is missing in the comment of these three patients is details on the doses and field sizes actually received, and co-morbidity which may have affected the outcome.

SUMMARY:

The results of the study showed that the F13 treatment showed a trend towards longer survival and longer time to distant recurrence, but that neither treatment showed an advantage in local control. Symptom control was achieved earlier in the F2 group.

CONCLUSIONS:

The authors conclude that neither treatment was clearly superior than the other, but that the F13 group had a marginal increase in survival. There was also a trend in better palliation in the F13 group with better pretreatment status. The

authors note that the time to recurrence in the F13 group was no different from that of the low dose group. They conclude from this that this means that the difference in survivals is due to the difference in the rate of distant metastases. They note the discrepancy in this result and that in the RTOG study of post-operative radiotherapy to a higher dose in which the radiotherapy had only an impact on the occurrence of local control, citing the reason for this to be the due to whether the disease has already metastasised or not.

This is reasonable and these conclusions are valid. The authors also go on to state that care in the interpretation of QOL measures from patient generated data must be used due to poor compliance, but in this study the clinician compliance was similar. There is no justification of the implied impression that clinician generated data is more accurate, and it is therefore valid to accept a clinician compliance rate of 70% but not a patient generated response rate. The authors also note that other studies have shown a marked discrepancy in the correlation between patient and clinician generated data, and yet make no attempt in this study to address this issue. The authors state "This trial clearly illustrates the relevance for clinicians of QOL endpoints...". Such a comment is generated from the authors opinions and not the trial results. The authors also present in their discussion a summary table percentage of the comparison of the two schedules. The place to present this is in the results section. The conclusions in the table are based on the trends observed as well as the results showing a statistical significance. The conclusions that a decision can be made in favour of one treatment over the other for the quality of life endpoints is not justified on the amount of data provided in the results. The only conclusion that can be made from the trial is that patients who received the F13 treatment had a non-significant longer survival and longer time to distant metastases, but that no treatment showed an advantage in local control.

PRESENTATION:

The presentation of this article is clear. The diagrams are appropriate and well labelled. The arguments are well presented.

ETHICS:

The study is ethically reasonable to do. Randomisation to two different treatment arms is ethical. I think the investigators are genuinely unbiased and set out not only to find out about the survival, but also to try to document the quality of life issues in these patients. It is recognised that quality of life assessment is difficult, and is a new field. No comment is made about ethics committee approval or about individual patient consent.

RECOMENDATIONS:

I would recommend that this paper be published. It addresses an important issue in a common disease and attempts to address quality of life issues which is also important. I would request the following corrections:

Methods:
The 15 patients with non-eligible histology should be excluded, or the authors should present results with and without these patients to show they made no impact on the results.

1. The last entry to Table 1 should be deleted, unless the authors can explain why this should be included.
2. Treatment deviations should be defined. More data should be given on the actual deviations which occurred, why treatments were delayed, what doses were achieved in these patients, and how fractionation was altered.
3. Authors should mention a mean and standard deviation of field sizes treated.
4. Reasons for the second course of radiotherapy need to be given.
5. Information of the cut-off points for the HAD scale as the accuracy is affected by this.
6. The handling of the palliation data could be clearer. The authors need to state how missing data was handled.

Quality Control:
Some review or verification of simulator and planning procedures needs to be included.

Statistics:

1. The authors should have calculated the new power when the study size was increased.
2. The palliation data should be presented as change in each patient’s score with time.
3. There should be some attempt to correlate the patient and physician scores.

Presentation:

1. More information or a brief review of the HAD and RSCL scales should have been included.
2. The presentation of the palliation data could be clearer, with the authors determining if the

results are the percentage of the results achieved, or the actual numbers.

SUMMARY:

This study tells me that both regimens are similar in achieving the palliation required and that either could be used. In the fitter patient I would be tempted to use the F13 regimen to give them the benefit of possible longer survival. It also tells me that the F2 regimen is just as efficacious in the patient in whom radiotherapy over a longer time will be difficult to give.

The study could have been split into 2 parts and those with metastatic disease could have been used to boost numbers in the palliation of symptoms arms, and those of good performance status could have been used to look at the survival issue. In doing this statistical power of the palliation arm could have been increased. It is reasonable to exclude those patients with metastatic disease from the survival arm as it is possible that higher doses to the primary will prevent metastasis beyond the chest thereby biasing the results if metastatic patients are included without being controlled for.

BIBLIOGRAPHY:

1. Medical Research Council Lung Cancer Working Party Bleehan N.M, Girling D.J, Machin D, Stephens R.J. A Medical Research Council (MRC) randomised trial of palliative radiotherapy with two fractions or a single fraction in patients with inoperable non-small cell lung cancer (NSCLC) and poor performance status. Br. J Cancer 1992;65:934-941.
2. Inoperable non-small-cell lung cancer (NSCLC):a Medical Research Council randomised trial of palliative radiotherapy with two fractions or ten fractions. Report to the Medical Research Council by its Lung Cancer Working Party. Br. J Cancer 1991;63:265-270.
3. Perez C.A., Brady L.W. Principles and Practise of Radiation Oncology. 1992 2nd Edn. J.B. Lippincott Company.
4. Peto R, Pike M.C, Armitage P, Breslow M.E, Cox D.R, Howard S.V, Mantel N, McPherson K, Peto J, Smith P.G. Design and Analysis of Randomised Clinical Trials Requiring prolonged observation of each patient. Br J Cancer 1976;34:585-612. Br J Cancer 1977;35:1-39.
5. de Haes J.C.J.M, van Knippenberg F.C.E, Neijt J.P. Measuring psychological and physical distress in cancer patients:structure and application of the Rotterdam Symptom Checklist. Br J Cancer 1990;62: 1034-1038.
6. Zigmond A.S, Snaith R.R. The Hospital Anxiety and Depression Scale. Acta Psychiatr Scand 1983;67:361-70.
7. Olschewski M, Schlugen G, Schumacher M, Altman D.G. Quality of life assessment in clinical cancer research. Br J Cancer 1994;70:1-5.
8. Gregory W.M, Bolland K, Whitehead J, Souhami R.L. Cautionary tales of survival analysis: conflicting analyses from a clinical trial in breast cancer. Br J Cancer 1997;76(4):551-558.
9. Daly L.E, Bourke G.J, McGilvray J. Interpretation and Uses of Medical Statistics. 4th Edn. 1996, Blackwell Science Ltd.