

Adjuvant Chemotherapy plus tamoxifen compared with tamoxifen alone for postmenopausal breast cancer : meta-analysis of quality adjusted survival.

Gelber R.D, Cole B.F, Goldhirsch A, Rose C, Fisher B, Osborne C.K, Boccardo F, Gray R, Gordon N.H, Bengtsson N, Sevelde P.

The Lancet 1996, 347:1066-1071.

AIM:

The aim of the paper is clearly stated under the heading of background as "to find out whether the benefit of adding chemotherapy to tamoxifen outweighs its cost in terms of toxic effects for postmenopausal patients. This is a reasonable question to ask. In the context of good palliation, the null result is still of value.

TRIAL SUMMARY:

This is a paper from the data of the Oxford overview database. The authors use the tool of meta-analysis to pool the results of several studies, and to look at the issue of quality of life. They then present a synthesis on the data to examine guidelines for the use of chemotherapy in palliative patients with breast cancer.

TRIAL DESIGN:

Patient Accrual

The authors have dredged the database of the EBCTCG overview (1) to look for studies which compared chemotherapy plus tamoxifen vs tamoxifen alone in patients with breast cancer. This resulted in 9 trials for their analysis. There was no attempt to update this database with new information (which was then 4 years old) and add new information to this.

Exclusions from the EBCTCG data included any study which was not a randomised trial, studies from Japan up until those beginning in 1985, and all studies from Russia. From the the Oxford overview there would be a minimum of 3332 patients who were directly involved in trials of chemotherapy plus tamoxifen vs tamoxifen alone, and possibly others in whom these were indirectly compared. It is not stated how many patients were eligible to be included and who were excluded for various reasons. This study may be subject to publication bias.

None of the study references cited in the bibliography which I were obtainable, included information on quality of life.(2-6) There are several possibilities as to how this information was obtained. One could assume that the authors had unpublished information from the authors of the original trials. The possibility exists that the authors used objective toxicity data which may or may not have been collected in a systematic way in lieu of true quality of life assessment. The validity of the latter method of collecting data and extrapolating it to represent patient perceived quality of life is to be questioned. Unless the authors can present some data to show that this is as accurate representation of quality of life the whole information on which this study is based is meaningless. The authors state in the methods section that the information was collected on an individual basis and included treatment, time to relapse, time to death or last follow. Nowhere is the collection of quality of life data mentioned.

The authors do refer to three papers regarding the Q-TWiST method. This includes Q-TWiST analysis of the Ludwig III trial (7) which is one of the trials analysed in this paper. It is possible that the same method was used to assign quality of life values for patients in the current study. The method includes data was collection for each patient of treatment time, time to relapse, whether local or systemic relapse, toxicity of treatments and overall survival. The times in each health state were then adjusted to account for recovery from toxicity eg. if a patient had alopecia following chemotherapy, the time for that patient on treatment was increased by three months to allow for regrowth of hair. For local recurrence in the mastectomy scar three months was added to relapse time to account for recovery from radiotherapy. Subjective toxic effects were graded by the investigators for severity and the time added to the toxicity time for that whole month. Details are in Table 2 of reference 7. In none of these methods was there allowance for permanent sequelae of treatment. Utility coefficients of 0.5 for relapse and toxicity and a value of 1 for Q-TWiST were assigned and the adjusted partitioned survival curves were calculated as described in Gelber et al. (8).

There are problems with the above procedure. While quality of life is certainly affected by the toxicity of treatment and relapse, the quality of life data is still based on objective rather than subjective measurements. There is no presentation that the assignment of utility coefficients of 0.5 for toxicity and relapse represents quality of life as perceived by the patients at that time. There is no allowance for change over time. On a realistic level, as patients become more unwell within a relapse category, there will be a gradual decline in their quality of life. There is no way that this method can handle the integration for several quality points over time, within a given health state. Patients quality of life can be affected by their mental state at the time, and for some patients this will fluctuate with time, as the disease evolves.

There is also the possibility that the values for quality of life coefficients was generated randomly using one of the three methods outlined in Belber et al (8).

Quality of Life measurement (QOL)

The above methods of defining quality of life are insufficient for a number of reasons. Quality of life instruments need to be purpose designed tools and need to be verified in the patient population in which they are to be applied. They need to have validity, reliability, and sensitivity i.e measure what they say they measure, be reliable, and be able to detect change(9). The scales need to be verified for the disease in question, across cultures, sexes, stage of disease, and across socioeconomic groups. If the data is not patient generated, then the objective assessment by the investigators needs to be shown to be valid for the patient group in question. This is usually achieved by the correlation of patient generated data and investigator generated data scores for individual patients. Investigator generated data is generally less reliable than patient generated data (9). A patients quality of life will change with time, and evaluation of a patients QOL over several time points is required. One of the statistical problems in the handling of longitudinal data from QOL studies is how to handle missing data, and how to handle variation over time.

Treatment Factors

The different treatment types are given in Table 1. All trials were of chemotherapy plus tamoxifen vs tamoxifen alone. The duration of chemotherapy ranges from 4 cycles to 24 cycles and the tamoxifen duration varied from 12 months to 60 months. While the different chemotherapy regimens are listed, the different drug doses are not, and so the reader is referred to the original papers to evaluate whether the chemotherapy regimens are comparable in terms of response and toxicity. Because it is likely that the authors have taken toxicity data from the original studies, and the analysis method uses toxicity weighted by time on treatment as a measure of quality of life, then this would be an important point. However, in this study the authors have had incomplete information as to the duration of treatment time and so have arbitrarily defined this time to be 6 months. The authors had information as to the number of cycles of treatment the different studies used, and it would be easy to obtain information regarding dosing intervals from the papers or the original researchers themselves.

Adjuvant treatments such as radiotherapy are not described. Radiotherapy can add to the toxicity of treatment and therefore reduce quality of life, thereby confounding the toxicity caused by the chemotherapy. There is no mention of control for other possible confounding factors such as surgical complications, and other co-morbidity.

Disease Factors

Patients were all node positive. There was variation in the receptor status in the patients and this is be clarified further. This is an omission as the response of breast cancer to tamoxifen is related to the receptor status. All patients were aged 50 years and over, meaning that the majority of women in the study would be likely to be receptor positive.

Followup

In the different studies, no mention is made of the followup arrangements in the groups. It is conceivable that in the chemotherapy patients who were being seen on a regular basis for treatment, that the data on quality of life was collected more frequently. In the event that the studies did prospectively collect and document valid quality of life data, there is still a potential confounder of the effect of more frequent doctor-patient followup on the analysis.

TRIAL CONDUCT:

The authors state that the data collection and quality control checks were as for the EBCTCG overview (1). There no further information in the original paper other than to mention that the data was checked exhaustively. There is no

further comment of the trial conduct and quality control thereafter. There should be further information of how non-completion of treatment was handled. The EBCTCG included all patients on intention to treat basis(1). There is no information on protocol deviations, or on patients lost to followup.

STATISTICAL DESIGN:

There is no calculation of projected power in the sample. Although the survival analysis is weighted, an estimate of this would be appropriate, with and without likely utility coefficient weightings.

Use of Collected data

There is no mention of how missing data in the original studies was handled, and whether the data was taken from the original studies as published, or if the authors made an attempt to address this in their own study. Excluded patients were not mentioned.

The Q-TWiST method is outlined in Gelber et al (8).

The method is outlined briefly:

Data is obtained for each patient for time in each health state. Mean time in each health state is calculated for each study. The method then assigns a value called a utility coefficient to represent the value placed on quality of life for each of the health states. Death is represented with a utility coefficient of 0 and disease free survival with 1.

The tamoxifen arm had a weighting of 1 for time in TOX. This was used as tamoxifen was common to both arms, but this is not strictly true. Utility coefficients of 0.5 were assigned arbitrarily to REL and TOX. If the method used in (7) are correct, the time in each health state has been altered to account for each episode of toxicity or relapse.

Meta-analysis

Mean time in each health state (with no standard errors or variance) are presented in Table 3 for each trial. This is a good measure to compare the groups, but is only valid if the treatments began immediately following diagnosis and surgery. It is these figures which are used in the meta-analysis. There is no measure of the heterogeneity across the studies, and so the reader is unsure whether it is applicable to group this data together (10-12), or even if it was valid to proceed with the meta-analysis. If heterogeneity occurs, the individual studies should be analysed more carefully, to find why there is a difference between the results, and if any differences can explain the observed result. Metaanalysis as a statistical exercise cannot correct for bad data.

The Q-TWiST method

The Q-TWiST method of analysing quality of life seems robust in principle. The statistical synthesis following this may not be. The limits of the Q-TWiST method include:

1. It assumes independence of utility coefficients for each patient for each state of disease, and for the entry point of the patient into the analysis.
2. It has no allowance for change in utility coefficient over time, as other factors in the patients' life change eg psychological state, financial dependence, functional status and progression of the disease. These factors may remarkably impair the patients quality of life, or their perception of their illness.
3. Q-TWiST assumes that the weightings can accurately be defined for each state. There is grave doubt that the authors in this study even attempted to look at quality of life, let alone that their values are accurate.
4. The method assumes that the quality adjusted time in each health state is proportional to the time spent in each health state.

Q-TWiST Synthesis and Meta-analysis

The authors present partitioned survival curves for their amalgamated data. The logic of partitioned survival curves is easily seen.

They go on to present a threshold utility (sensitivity) analysis for their data as a method to compare the two treatments, and a Q-TWiST gain function as a measure of the change in the benefit from the chemotherapy treatment arm with time. The methods are described in Gelber et al(8).

The threshold analysis refers to the relative value of gain in quality adjusted survival adjusted for the toxicity if the treatment. Confidence intervals are calculated for this.

The authors have then entered the mean survivals into a regression analysis. The authors do not state which models were used to combine the results of the individual studies. They then used a generalised least squares regression to look at the variation across the studies and to derive weightings to apply to the trials. Following this they performed a separate regression analysis for each of the trials and used this to generate data for each health state. They then looked at the variation between trials, and calculated values for Q-TWiST. All possible values for utility coefficients were then derived from the equation and entered into a threshold utility analysis.

There is considerable debate regarding weighting of trials (10-12,14)). Weightings were assigned to reflect the quality of the study, and to make allowance for the different followup times. The authors present the results from the individual trial data in Table 3 as mean time in each health state. These have been adjusted for utility coefficients of 0 for death, 1 for DFS and Q-TWiST and 0.5 for TOX and REL. Unfortunately there is no measure of the dispersion around the mean. Following this Table 4 presents the combined time in each health state from all studies of (combined therapy-tamoxifen alone) with the 95% confidence interval. It would be more informative to either give the standard error of the mean as this is a measure of how good the mean is as a representative of the population mean. Partitioned survival curves are accurately depicted, as are the graphs of threshold utility and Q-TWiST gain function.

The values of dispersion from the original trials are lost in the initial regression analysis. There is no mention of how standard errors and variances would be handled in the analysis, or whether they are totally ignored. Further, the authors provide no documentation of the error involved in the entire process, as each step of the process will have associated with it, assumptions and also an error. The error will compound with each step of the process.

The validity of the entire process has not been addressed. The statistical methods used are robust in their own right, and the Q-TWiST analysis, despite its temporal limitations would appear simple enough in concept, but how well this is able to be applied to the patient population as a predictive tool as the authors suggest is in doubt. This is especially so given that the authors leave some doubt as to how the "quality of life" data was derived.

CONCLUSIONS:

The authors conclude that within the 7 years of followup, there was no improvement in Q-TWiST for any weightings of toxicity or relapse for the comparison of chemotherapy and tamoxifen vs tamoxifen.

In their discussion, the authors address the issues of applicability of the method to medical decision making. They make no attempt to address the shortfalls in the study, or to discuss the applicability of the methods which are unverified in a patient population.

PRESENTATION:

The presentation of this difficult topic to the non-statistician is difficult to understand. As a paper the methods described are not unambiguous enough for the reader to understand exactly what has been done, and the exact methods used to perform the regression analysis. It may have been more appropriate to provide a flow diagram of the steps of the regression analysis, and also to provide a brief outline of the rationale behind the different statistical procedures involved. The method of primary data collection is ambiguous. The rationale for performing the study is clearly stated and laudable. Illustrations are appropriate. More data could have been given on the original studies, and exactly what data was taken from the trials as published, and what was sought as inside information from the authors..

ETHICS:

In meta-analysis it is impossible to seek individual patient consent. The authors approach their topic from an unbiased point of view, and the study question is ethical. The authors have presented a method with which to make decisions regarding patient treatment without verifying the method based on true patient data. This is ethically wrong.

RECOMMENDATIONS:

I would suggest that this paper not be published in its present form. I would insist on the following changes prior to publication:

1. The authors should be more specific on how the quality of life data was obtained, and what specifically was obtained from the raw data.
2. The authors should comment on how missing data, incomplete data was handled.

3. The authors should comment on eligibility criteria, how many studies were initially considered, and how many were excluded and for what reasons.
4. Authors should comment on the lack of updates, if there was any attempt to search for recent studies, and what methods were used.
5. The authors should be more specific about the statistical analysis and state if the assumptions required to perform a valid analysis were verified.
6. More data should be given on the handling of prognostic variables in the regression analysis, or if this was left purely to the randomisation procedure in the initial trials.
7. The data and models generated with this data set needs to be verified against true patient generated data. If this is not done, the instruments which the authors propose for medical decision making (the threshold utility analysis) is useless.
8. Authors should be more specific about the exact regression methods used. It would be useful to present the method of analysis in a flow diagram form. More information about the rationale of the different parts of the statistical analysis for the non-statistician would improve the readability of this paper.

In addition to the above, the authors may like to explore in their discussion the limitations of the model presented, and would be encouraged to comment on why the accelerated failure time regression method was not used (9). This method has been developed by the authors and others and would appear to be a more realistic way of modelling this data as it takes into account previous health states and previous transitions.

SUMMARY:

This paper used data from the EBCTCG database in an unspecified way to generate quality of life data to be used in the prediction of benefit to be gained by chemotherapy vs tamoxifen comparisons. The method is ambiguously described and their conclusions are based on objectively derived data. There is no correlation with actual patient generated data. While this work provides interesting avenues for further research, there are inherent difficulties in assessing patient QOL data in this way, and making treatment decisions based on the results achieved in this method.

BIBLIOGRAPHY:

1. Early Breast Cancer Trialists' Collaborative Group. Systemic treatment of early breast cancer by hormonal, cytotoxic or immune therapy. *Lancet* 1992;339:1-15, 71-85.
2. Overgaard M, Christensen J.J, Johansen H, et al. Evaluation of radiotherapy in high risk breast cancer patients:report from the Danish Breast Cancer Cooperative Group (DBCG82) Trial. *Int J Radiat Oncol Biol Phys* 1990;19:1121-24.
3. Fisher B, Redmond C, Legault-Poisson S, et al. Postoperative chemotherapy and tamoxifen compared with tamoxifen alone in the treatment of positive node breast cancer patients aged 50 years and older with tumours responsive to tamoxifen:results of the National Surgical Adjuvant Breast and Bowel Project B-16. *J Clin Oncol* 1990;8:1005-1018.
4. Senanayake F. Adjuvant hormonal chemotherapy in early breast cancer:early results from a controlled trial. *Lancet* 1984;ii:1148.
5. Rivkin S.E, Green S, Metch V et al. Adjuvant CMFVP versus tamoxifen versus concurrent CMFVP and tamoxifen for postmenopausal node-positive breast cancer patients:a Southwest Oncology Group study. *J Clin Oncol* 1994;12:2078-85.
6. Boccardo F, Rubagotti A, Bruzzi P et al, for the Breast Cancer Adjuvant Chemo-Hormone Therapy Cooperative Group. Chemotherapy versus tamoxifen versus chemotherapy plus tamoxifen in node-positive, oestrogen receptor-positive breast cancer patients:results of a multicentric Italian study. *J Clin Oncol* 1990;8:1310-20.
7. Goldhirsch A, Gelber R.D, Simes RJ, Glasziou P, Coates A.S. Costs and benefits of adjuvant therapy in Breast cancer:A Quality adjusted Survival Analysis. *J Clin Oncol* 1989;7(1):36-44.
8. Gelber R.D, Cole B.F, Gelber S, Goldhirsch A. Comparing treatments using Quality adjusted Survival:The Q-TWiST Method. *Am Stat* 1995;49(2):161-169.
9. Olchewski M, Schlugen G, Schumacher M and Altman D.G, Quality of life assessment in Cancer research. *Br J Cancer* 1995;70:1-5.
10. Mosteller F and Colditz G.A. Understanding Research synthesis (Meta-analysis) *Ann Rev Public Health* 1996;17:1-23.
11. Greenland S. Invited commentary:a Critical Look at some Popular Meta-analytic methods. *Am J Epidem.* 1994;140(3): 290-299.
12. Shapiro S.Meta-analysis/Shmeta-analysis 1994;140(9):771-778.
13. Katz M.H,Hauck W.W. Proportional Hazards (Cox) Regression. *J Gen Int Med* 1993;8:702-711.
14. Greenland S Quality scores are useless and Potentially Misleading. *Am J Epidem.* 11994; 140(3):300-301.