Paper 259-25

DESIGN AND ANALYSIS OF CANCER CONTROL STUDIES USING THE SAS SYSTEM

Jeff A. Sloan, Paul J. Novotny, Charles L. Loprinzi
Mayo Clinic, Rochester, MN 55905

ABSTRACT

This paper focuses on the details of study design, data elements and analytical methods necessary for oncology clinical trials when the primary endpoint involves symptom control rather than treatment outcome. We will present work carried out on previous North Central Cancer Treatment Group (NCCTG) Cancer Control (CC) clinical trials which involved standard and novel results which were obtainable via SAS code. Measurement issues with respect to pain, stomatitis, and other quality of life related endpoints will be highlighted. Considerations involved in determining clinical significance, and related power calculations, for these types of endpoints will be presented.

INTRODUCTION

A keystone article in the New England Journal of Medicine in 1986 by Ian Tannock indicated that since a cure for cancer was not on the immediate horizon, perhaps it was "time to treat the patent, not the disease". This coincided with an upsurge in the amount of attention paid to cancer symptom management and control issues. Trials targeted at survival and tumor response began to be supplemented by studies targeted to other endpoints concomitant with the patient's cancer experience. The primary goal of such studies was controlling patient suffering rather than curing the cancer.

The evolution of cancer control research brought with it some unique challenges in the design and analysis of clinical studies. Issues such as deciding upon a primary endpoint, choosing a measurement approach, defining clinical significance and dealing with missing data became the focus of this work. Along the way, new statistical methodology was necessary to handle the peculiarities of analyzing these new endpoints and designs.

In this paper, we explore the uniqueness of cancer control research and present some ways that we have found the SAS system particularly useful. A primary role for SAS is, of course, the implementation of alternative analytical approaches in a series of SAS macros and related programming gymnastics. The versatility of the SAS system was demonstrated in the development of secondary roles of producing routine summary reports for regulatory requirements and handling monitoring tasks. We demonstrate below the results of this work using completed Mayo Clinic Comprehensive Cancer Center and NCCTG oncology clinical trials. A cornucopia of SAS procedures and macros were involved. The ultimate goal is to relate our experiences, provide some guidance for others facing similar challenges in oncology clinical trials and, in the best tradition of SUGI, share some code.

DESIGN

The design of any clinical trial is a multi-faceted and complex process. Cancer control studies add unique challenges over and above standard phase II/III oncology clinical trials. Although there are measurement issues in studies assessing tumor response and survival, the issues have certain consistency across trials. In cancer control studies it is possible that the basic endpoints and designs will change over a series of clinical trials. Knowledge often is gained regarding the measurement, distribution, and useful analytical methods for moving forward the science of cancer control.

Endpoints

The vagaries of designing cancer control studies relative to more standard treatment trials begins with the identification and definition of the endpoints under study. It is difficult to measure directly the entities that represent subjective or intangible clinical phenomena. Many intangible constructs, such as pain, for example, are easily recognizable subjectively but are difficult to quantify objectively. We all know what pain is, but we are unable to directly observe and measure pain in the same consistent and reliable fashion that we measure tumor response. We are left, as a result, with having to use surrogate measures of the construct under study. For example, although the research on pain measurement is large and impressive in its scope and science (Cleeland et al, 1994), at the end of the day we are left with measuring pain by asking patients to give us a number between 1 and 10 supplemented possibly by descriptive adjectives.

At the heart of this type of measurement research is the assumption that improvements in the intangible constructs will coincide with, or at least relate to, the data which can be obtained from the patient or clinical observer. These issues are referred to as reliability and validity in classical measurement literature and are beyond the scope of this work. Readers are referred to a tome of collected literature on this topic for further reading (Spilker, 1996).

The choice of endpoints can be a difficult one, even in a seemingly simple situation. For example, patients with advanced cancer often experience a "wasting away" or "crashing" wherein they experience rapid and substantive weight loss leading to death. This situation, known as anorexia/cachexia, imparts great suffering on patients and family members. Studies have been initiated to stem the anorexia/cachexia syndrome by investigating ways in which patients can at least maintain weight in the presence of an onerous tumor burden. It would seem to be a simple matter of assessing weight gain and finding agents that will increase a patient's weight, but nothing is ever that simple (Cella, 1993). The mere gaining of weight may in fact be an accumulation of fluid which is not an indication of a beneficial result for a patient with advanced cancer (Burman & Chaberlain, 1996). One could consider assessing muscle mass and tone but such assessments typically involve rather invasive procedures, which may be ill advised for many patients with advanced cancer. Appetite stimulation is a potential surrogate for weight gain, although there is no guarantee that an enhanced appetite will translate into improved weight. Adding to the challenge is finding a means by which to assess, in a reliable and valid manner, the degree of appetite a patient is actually experiencing. The degree of nausea and vomiting are other concomitant intervening variables that can express the relative success of an agent targeted at anorexia/cachexia. The ultimate approach we have taken with such work is to treat the target as a multivariate endpoint. Our studies have assessed all of the above endpoints and have utilized contemporary multivariate procedures implemented in the SAS system for analysis.

Postmastectomy breast cancer patients often experience periods of augmented hot flash activity and lymphadema. Both symptoms are at the least bothersome and at the worst debilitating. Choosing the endpoint to target for both hot flash activity and lymphadema is not a simple task. We have used the approach of a bivariate endpoint reflecting the frequency and intensity of hot flashes and designing the studies accordingly. For lymphadema, the task is even more daunting due to the multivariate nature of the condition. Table 1 indicates the long list of endpoints that are obtained for such studies, including limb volumetric and circumference measurements, subjective assessments of swelling and feelings of pressure and so on. We chose to use a geometrical combination of these values combined into a single endpoint that could be displayed graphically (Figure 1). The Data Set Graphical Interface (DSGI) in SAS allowed for the construction of this novel graphic which ultimately replaced all of the analytical p-value results for the given study (Loprinzi et al, 1999). The images displayed are the mean values for the swollen limbs superimposed on data for the unaffected arms for patients at baseline and at six and twelve months post treatment initiation in a classic two-period crossover design. While some may argue whether the final product looks more like carrots or rockets than human limbs, the six images portrayed are virtually identical in shape, size and relative proportion of the swollen limb to the unaffected arm.

Chemoprevention studies form a subset of cancer control research and falling under this subset are studies aimed at preventing lung cancer by targeting smoking cessation. This is a particularly interesting group of people to study because they are radically different from cancer patients. It has been suggested that smokers trying to quit exhibit behavior similar to hardened drug addicts and should be treated with the same degree of caution. Our experience confirms this suggestion. While cancer patients are occasionally non-compliant, it is typically due to illness or undue toxicity from the study treatment. As many as half of smokers in a clinical study, however, are routinely non-compliant and may exhibit challenging behaviors. Asking a smoker whether they are smoke free may not be sufficient to obtain a reliable estimate of smoking cessation treatments. Chemically verified smoking abstinence is hence obtained by the use of a carbon monoxide monitor. If the smoker refuses to give such data or fails to appear for appointments, then they are classified as a smoker for the purposes of analysis in an intent to treat fashion.

A final point is some advice on how to handle multiple endpoints in a single study. There are standard multivariate procedures available in SAS through PROC MANOVA and other routines. A useful approach is to analyze each endpoint using univariate procedures and supplement these results by multivariate processes. We have implemented in SAS a general method for clinical trials with multiple endpoints which was first proposed by O'Brien (1984). This allows for a single p-value to express the degree of difference in the relative rankings of observations from treatment groups across an array of endpoints.

Quality of Life Assessment

A major target for cancer control studies is the assessment of patient quality of life (QOL). QOL has been demonstrated to be prognostic for patient survival (Degner & Sloan, 1995; Sloan et al, 1998a). A sense of overall well-being can be captured by the use of simple QOL measurement tools (Aaronson, 1991; Cella et al, 1996, Cox, 1992). QOL is, however, a multi-faceted and multivariate entity comprised of numerous constructs (Spilker, 1996). Unfortunately, there remains a substantial amount of controversy regarding the accuracy and clinical significance of measuring QOL (Laura-Munoz & Feinstein, 1999; Leplege and Hunt, 1997).

Several issues must be considered in choosing QOL endpoints for cancer control studies (Fayers et al, 1997; Moinpour et al, 1991; Sloan et al, 1998c). A useful recipe is to a priori assess which elements of QOL are of interest and likely to be affected by the treatment under consideration. Once that list has been constructed, it is then suggested to peruse the existing tools to see which, if any, contain the majority of the constructs of interest. There is a tendency for researchers to take the path of least resistance and grab a tool "off the shelf" for inclusion into the study. Such a practice heightens the likelihood that no impact will be observed on patient QOL by the mere fact of including a number of constructs that are irrelevant (Sloan et al, 1998). For example, the SF-36 is an often–used and one of the most validated instruments for measuring patient QOL. The tool, however, was designed for general patient and healthy populations. The tool includes questions of physical functioning such as whether or not a patient can climb stairs and carry groceries. Asking such questions of advanced cancer patients might be meaningless. What might happen is that none of the patients will be able to do these things at baseline and none will be able to do them throughout the course of the study. Hence this will add items of zero variance to the QOL evaluation which will make the analytical results less sensitive to change and bias the study towards an inflated type II error rate (Sloan et al, 1998b).

It is recommended, therefore, is to use a standardized QOL instrument which covers most of the constructs of interest and supplement this by individual, study specific questions. One may consider drawing individual items from existing tools, but this must be done with great care to retain the integrity of the psychometric properties of the original tool.

We believe that in the presence of substantial barriers such as have been delineated above, it is important to observe the KISS (keep it simple and straightforward) principle in designing and analyzing cancer control trials. As such, we incorporate measures that are simple, brief and are targeted at the global constructs of quality of life that are likely to change over the course of the trial (Sloan et al, 1998a-c). Our objective is to get global measures having these three constructs in every trial in an efficient manner without causing undue burden to the patients.

CLINICAL SIGNIFICANCE

Clinical significance is an important characteristic of cancer control clinical trials. There are several ways of assessing clinical significance a priori in the design phase although there is not a generally accepted optimal approach. Sometimes the literature provides guidance for determining clinical significance. For example, supplementary analysis of raw pain data can be undertaken by using a classification system for pain scores of 0-3, 4-6 and 7-10 as representing mild moderate and severe pain respectively.

A direct approach is to ask the patient if their QOL has been impacted by the treatment trial. Other work involves asking patients whether their QOL has changed or not and relating this categorization to changes in the average QOL scores over time so that a minimally observable difference may be determined (Osoba, 1999). A further approach involves initial estimation of the QOL standard deviation by way of the empirical rule of statistical theory with the classification of effect sizes due to Cohen (1988) so that a priori one can decide sample size based upon the idea of whether a small, moderate or large effect size may be detectable (Sloan et al, 1997, 1998b,c).

A comparison of these methods indicates that all approaches give similar answers (Symonds and Sloan, pending). The minimally observable difference approaches and the empirical rule effect size method all indicate that a shift of 10 points on a 0-100 scale for a QOL measure is likely clinically significant. A number of tool developers have suggested such a benchmark for declaring a change in distribution has occurred. We have used this in several clinical trials as a supplementary QOL endpoint by comparing the

proportion of patients who achieve a clinically significant shift of ten points over the study period.

Consider the examples for reducing hot flash activity discussed earlier. In hot flash activity, we can typically reduce the number of hot flashes an average of one hot flash per day and the hot flash score by between 25% and 30% by administering a placebo preparation (Figure 2). We use this knowledge to power hot flash studies so that we can detect differences of at least this magnitude (Loprinzi et al, 1998; Quella et al, 1998; Barton et al, 1998). In particular, while investigating the efficacy of vitamin E for reducing hot flash activity, we observed that women reported a reduction from roughly six hot flashes per day to around 4.7 hot flashes per day through the use of a placebo. People receiving vitamin E reported a reduction to 4.0 hot flashes per day (Loprinzi et al, 1998). The study was powered to detect a difference of 1 hot flash per day between the placebo and vitamin E treatment groups. The results were on the cusp of statistical significance, producing p-values of between 0.55 and 0.45, depending upon the statistical procedure utilized (Figure 3). More important than statistical significance was the fact clinically a reduction of 0.7 hot flashes per day may or may not be important. The study results provided a unique opportunity to recommend vitamin E on the grounds that it may produce the placebo effect at least and perhaps a bit more. Since vitamin E is inexpensive, non-toxic, and readily available, we were able to recommend it on the grounds of clinical rather than statistical significance.

Whichever power analysis approach for determining clinical significance is used, the most important step is to decide up front the benchmarks that will be considered as evidence that QOL has been significantly altered. This is a priority area of research in need of further development.

Stratification

Stratifying a randomized study by potentially confounding concomitant variables is an idea not restricted to cancer control studies. The nature of the endpoints, however, typically necessitates a need for further stratification due to the inherent measurement error of the endpoints themselves. This raises the important issue of the number of stratification factors that is feasible in a given study. Apart from logistic considerations, a suggested upper limit for the number of stratification factors has been proposed as being one half of the number of observations per treatment group (i.e., n/2, Therneau, 1993). Treatment assignment itself is often carried out by a method of randomly permuted incomplete blocks or by using a dynamic allocation procedure which balances the marginal distributions of the stratification factors between the two treatment-sequence groups (Pocock and Simon, 1975).

Studies involving advanced cancer patients are in need of stratification factors to control for the illness severity being experienced by the patients. Clearly if one group of patients is in poorer shape than another, it will bias the results of the trial in favor of the other treatment. So how does one assess the relative liveliness of a patient with advanced cancer?

One method is to ask the attending physician. We have demonstrated that a physician can provide reasonable estimates of expected remaining lifetime for patients in their care (Loprinzi et al, 2000). Figure 4 displays survival curves for a sample of advanced cancer patients stratified by physician estimate of expected survival classified into an ordinal scale of less than four months, four months through six months, and more than six months. Despite the simplistic nature of the variable, the curves separate substantially among the three groups.

Expansion of this work included a well known indicator of physical well-being known as the Eastern Co-operative Oncology Group (ECOG) performance status measure which classifies patients along a four point ordinal scale in terms of their ambulatory ability. We undertook an exhaustive search for variables that would supplement this often-used stratification factor involving over ten years of research and 2,000 patients. The end result of this search was the construction of a simple index for stratifying patients with advanced cancer into those with good, bad, and uncertain prognosis. This GBU index (Sloan et al, submitted) includes four variables: the ECOG performance status, the physician's estimate of survival described earlier, as well as subjective measures of the patient's appetite and physical status. Figure 5 demonstrates that the GBU index can differentiate patients with advanced cancer into discernibly separate categories and hence improve the efficiency of the statistical design.

MONITORING

A couple of particular studies caused us to use the SAS system in a somewhat non-standard way. First, contractual obligations required standardized monthly reports for a particular cancer control chemoprevention trial. Using null datasets, PROC PRINTTO, and PROC TABULATE, we produced such reports that could be generated automatically and serve as audit instruments (Figure 6). These reports actually became the focus of contract negotiations because of their central place in the analytic process of the trial. A second trial required summary reports to be generated for each individual patient. PROC SQL and null datasets allowed for speedy data entry, retrieval, and reporting.

A final note in the use of the SAS system for monitoring clinical trials. There have been a number of SAS macros produced over the years to generate "event charts" so that the trajectory of each patient through the clinical trial process can be displayed parsimoniously. While these programs are typically protocol specific in cancer control studies, due to the different nature of the events one needs to track for each study, the use of this technique in SAS is invaluable.

ANALYTICAL APPROACHES

There is, at present, no generally accepted statistical approach for cancer control data that is considered to be optimal (Sloan et al, 1999a-c). As such, analysis of the data for cancer control studies is best carried out in a number of complementary ways in the form of a sensitivity analysis. If the results replicate across a number of statistical approaches then you can have confidence that the findings are not a function of the assumptions underlying the statistical procedures.

Statistical methods used on such data have often included high power statistical procedures such as repeated measures analysis of variance (ANOVA/GEE), polynomial effects models and other multivariate analyses (Cella et al, 1995, for example). Our approach has been to use simple straightforward procedures and hold the more complicated analyses in abeyance until the basic questions have been answered. The assumptions for the more involved procedures are both numerous and complex. Further the assumptions are not tenable for use in measuring intangible constructs with inherent measurement error, a lack of normality and a measurement level somewhere between ordinal and interval level data. We demonstrate the ability of the SAS system to handle these challenges using a series of examples.

One of the greatest advantages of SAS is that one can gather together, in a single macro, a wide variety of analyses that can form the core results for a particular study type. For example, the hot flash studies described previously amounted to a series of eight different clinical trials, all involving two-period, two-group crossover designs. We collected together various statistical procedures used on such designs and implemented them in a series of SAS macros (Sloan et al, 1997). We updated this work by adding a Bayesian Markov Chain Monte Carlo (MCMC) analysis to this series of routines (Mandrekar et al, 1999). Space limits our ability to describe the procedures in detail, but suffice it to say that the SAS macros run through the standard analytical methods for crossover designs and then implement some alternative models and graphical representations.

Missing Data: Valid and reliable QOL data can be collected from cancer patients as demonstrated by numerous clinical trials. The reality of working with seriously ill patients is that some will die during the course of the trial. There are several ways to handle this eventuality in the design and analysis phase of the trial.

One can choose among several methods for adjusting for missing data (Fairclough, 1997). Recent work in the area of QOL research has focused on multiple imputation methods. The number of alternative approaches here is staggering. One can choose to impute missing data by carrying forward, the last score obtained, the minimum score obtained, the maximum score obtained, the average value obtained or a zero value to reflect the fact that the patient is no longer living. Each approach has application.

Missing data can be handled to a certain degree in the analysis of a clinical trial. The primary goal is to first ascertain if there are any reasons to expect that the data missing are absent due to a random or systematic force.

In advanced cancer patients, it is often reasonable to assume that people do not provide data because they are too ill to participate further in the trial. One can suggest imputing missing data accordingly. For example, if we are examining patient QOL in such a study, it is reasonable to assume that patients with missing data have QOL no higher than their last provided observation. Alternatively, one might assume that their QOL has dropped to the point of being non-measurably different from zero. Imputing data using

both assumptions provides a form of sensitivity analysis in that they reflect the best and worst case scenarios (Figure 7). Alternatively there are a number of imputation methods that suggest using the average value carried forward (AVCF), or some other functional. Figure 7 displays average QOL for a study of patents with advanced cancer receiving hydrazine sulfate. One can see a considerable amount of attrition from 108 patients at baseline to only two patients providing data at twelve months. It can be seen that the nature of imputation method used in this extreme example provides markedly different QOL profiles over time. In particular the line marked (OA), representing using only the available data at each point seems to present the best case for high QOL in this patient population because the only folks who are left are ones with good QOL. Similarly, if we adjust this profile by scoring all missing data as zero, we see that the attrition in the study suggests that QOL for the original 108 patients drops sharply and consistently towards zero over the course of a year. The most important aspect of such a sensitivity analysis is to assess that the imputation methods do not alter the basic conclusion of group comparison. In our example, Figure 8 demonstrates that whether the OA or zero carried forward imputation method is used, the treatment comparison between placebo (PL) and hydrazine sulfate (HS) produces the same finding that QOL is the same for both groups.

Multiple imputation methods are another popular approach. Unfortunately, such methods are assumption heavy and once again are contradictory to the KISS principle. Unless more than 20% of the data are missing, the results one gets from any imputation approach will likely be the same (Fairclough, 1997).
One of the analytical strengths of SAS, its graphical capabilites, has been often maligned as being difficult to master. We have found that typically people who make such claims have not taken the time to explore its facilities or are content with point and click capabilities to produce standard graphics. The SAS system provides the flexibility to produce novel graphical illustrations with relatively little effort. SUGI has been marvellous for containing talks demonstrating graphical gymnastics and so we will not spend time demarcating the long list of accomplishments in this area. The reader may obtain excellent examples in the works of Michael Friendly both in text and on the web (Friendly, 1994). We close this section with two brief examples of how simple graphics provided primary results similar to the coumarin example cited earlier.

The first example (Figure 9) has been dubbed internally as either the “spider”, “two bugs crashing”, or the ungodly mess. We prefer to call it the mirror image stream plot and it is a useful tool for identifying treatment differences in the complete distribution of observed data for a given study. Pictured in Figure 9 are the individual weight statistics for each patient over time on one of the anorexia/cachexia studies described earlier. The horizontal reference line indicates stable weight. The x-axis is a mirror image so that weights for one treatment group (megesterol acetate) are portrayed to the left of the vertical reference line and weights for another treatment group (fluoxetine) are displayed to the right. If the two treatment groups were equivalent, we should see a mirror image in a symmetric display on either side of the vertical reference line. This graphic has an advantage over simply displaying average values in that the proportion of patients who actually benefit from the treatment can be observed. Figure 9 shows that more patients on megesterol acetate achieve weight gains than do those receiving fluoxetine. The variability of the picture also indicates that megesterol acetate is not a uniformly successful agent for inducing weight gain for advanced cancer patients.

Finally, we relate another graphic easily constructed in SAS for meta-analysis. Known as a Forrest plot, Figure 10 presents the results of a study into a seemingly differential in mouth soreness reported by men and women undergoing standard chemotherapy treatment. This work evolved out of a simple observation for NCCTG study 959251, which, as demonstrated in Figure 10, revealed a statistically significant difference in that women reported a greater amount of mouth soreness. This was a surprising result because it was the fifth clinical trial examining alternative agents to reduce mouth soreness in chemotherapy patients. None of the previous studies had demonstrated a significant difference between men and women as is verified in figure 10. The studies had not been powered to detect such differences however. The series of data shows a consistent effect with women reporting greater mouth soreness. The meta-analytic p-value based on a total of over 700 patients is statistically significant. This meta-analysis have been confirmed by a further study involving over 3,000 patients (Sloan et al, in press).

SUMMARY

We hope that this paper gives the reader an appreciation for the richness and variety of challenges

faced in cancer control research. It is both the appeal and source of frustration. SAS code and/or reprints of the studies discussed may be obtained from the authors via email at jsloan@mayo.edu or by contacting:

Jeff A. Sloan, Ph.D.
Mayo Clinic
Charlton 6, 200 First Street SW
Rochester MN 55905
(507)284-9985.

References

Aaronson NK. Methodological Issues in Assessing the Quality of Life of Cancer Patients. Cancer 67 Suppl.: 844-850, 1991.

Barton DL, Loprinzi CL, Quella SK, Sloan JA et al. Prospective evaluation of vitamin E for hot flashes in breast cancer survivors. JCO 16: 495-500,1998.

Burman R Chamberlain J. The assessment of the nutritional status, caloric intake, and appetite of patients with advanced cancer. pp. 83-93 in Cachexia-Anorexia in Cancer Patients. Oxford University Press, London , 1996.

Cella DF Bonomi AE et al. Quality of life and nutritional well-being: measurement and relationship. Oncology 7: 105-111, 1993.

Cella, D.F. Quality of life outcomes: measurement and validation. Oncology 10: 233-246, 1996.

Cleeland CS, Ryan KM. Pain assessment: global use of the brief pain inventory. Ann Acad Med Singapore 23: 129-138, 1994.

Cohen J. Statistical Power Analysis ofor the Behavioral Sciences. Lawrence Erlbaum Associates Hillsdale New Jersey, 1988.

Cox DR, Fitzpatrick R. Quality-of-life Assessment: Can we keep it simple? J R Statis Soc 155: 353-393, 1992.

Degner LF, Sloan JA. Symptom Distress in Newly Diagnosed Ambulatory Cancer Patients and as a Predictor of Survival in Lung Cancer. J Pain and Symptom Management 10: 423-431, 1995.

Fairclough, D.L. Summary Measures and Statistics for Comparison of Quality of Life in Clinical Trial of Cancer Therapy. Stat in Med 16: 1197-1209, 1997.

Fayers, P.M., Hopwood, P., Harvey, A., et al. Quality of Life Assessment in Clinical Trials – Guidelines and a Checklist for Protocol Writers: the U.K. Medical Research Council Experience. Europe J Cancer 33: 20-28, 1997.

Friendly M. SAS for Statistical Graphics. SAS Institute, Cary, NC, 1994.

Lara-Munoz C, Feinstein AR. How should quality of life be measured? Journal of Investigative Medicine 47(1):17-24, 1999.

Leplege A., Hunt S. The problem of quality of life in medicine. JAMA 278(1): 47-50, 1997.

Loprinzi CL, Kugler JW, Sloan JA, et al. Evaluation of coumarin for treatment of postmastectomy lymphadema. NEJM 340: 346-350, 1999.

Loprinzi CL, Sloan JA, Quella S et al. A phase III comparison of megesterol acetate versus dexamethasone versus fluoxymesterone for the treatment of cancer anorexia/cachexia. JCO, 2000 .

Mandrekar J, Sargent DJ, Novotny PJ, Sloan JA. A general Gibbs sampling algorithm for analyzing linear models using the SAS system. Proceedings of the SUGI 24, 1540-1545, 1999.

Moinpour, C.M., Feigl, P., Metch, B., Hayden, K., Meyskens, F.L., and Crowley, J. Quality of life end points in cancer clinical trials: Review and Recommendations. J Nat Can Inst 81: 485-495, 1991.

Salant I, Dillman D. Market Surveys: The Total Design Method. John Wiley & Sons, New York 1996.

O'Brien PC. Procedures for comparing samples with multiple endpoints. Biometrics 40: 1079-1087, 1984.

Osoba D, Rodrigues G, Myles J, Zee , Pater J. Interpreting the significance of changes in health-related quality of life scores. J Clin Oncol 16: 139-144, 1998.

Pocock SL, Simon R. Sequential treatment assignment with balancing for prognostic factors in control clinical trials. Biometrics 31: 1103-1115, 1975.

Sloan JA, Novotny PJ, Loprinzi CL, Nair, S. Graphical and analytical tools for the analysis of two-period crossover clinical trials. Proceedings of SUGI 22, 1312-1317, 1997.

Sloan JA Loprinzi CL Kuross SA Miser AW et al. Randomized comparison of four tools measuring overall quality of life in patients with advanced cancer. JCO 16: 3662-3673, 1998a.

Sloan JA O'Fallon JR Suman VJ Sargent DJ. Incorporating quality of life measurement in oncology clinical trials. Proc Biometrics section of the American Statistical Association, 282-287, 1998b.

Sloan JA, Novotny PJ, Loprini CL. Analyzing quality of life endpoints in clinical trials via the SAS system. Proceedings of SUGI 23, 1213-1218, 1998c.

Sloan JA, Novotny PJ, Loprinzi CL. A simple stratification factor for survival in advanced cancer patients. JCO 17, 2000 (in press).

Spilker, B.(ed.) Quality of Life and Pharmacoeconomics in Clinical Trials. New York, NY: Lippincott Raven, 1996.

Symonds T, Sloan JA. Health-related quality of life measurement in clinical trials: when does a statistically significant change become clinically relevant? (unpublished manuscript).

Therneau, T. How many stratification factors is too many to use in a randomisation plan? Controlled Clinical trials 14: 98-108, 1993.

Table 1: Endpoints for Lymphedema

Circumference of Affected Hand Circumference of Affected Wrist Circumference of Affected Arm at 30 cm Circumference of Affected Arm at 40 cm Circumference of Affected Arm at 50 cm
Ratio of Circumference of Affected to Normal Hand Ratio of Circumference of Affected to Normal Wrist Ratio of Circumference of Affected to Normal Arm at 30 cm Ratio of Circumference of Affected to Normal Arm at 40 cm Ratio of Circumference of Affected to Normal Arm at 50 cm
Distal Edema: ratio of the sum of the circumferences at hand, wrist, and 30 cm divided by the corresponding sum on the normal arm
Total Edema: ratio of the sum of all circumferences on the affected arm divided by the sum on the normal arm
Volume of Affected Arm estimated from the circumference measurements
Volume of Affected Arm Divided by the Volume of Normal Arm
Patient Reported pressure pain in arm, heaviness in arm, arm tightness, loss of arm mobility, arm swelling
Patient preference of which crossover period they preferred
Patient Rating of whether they felt the tablets were helping

Figure 1: Arm Means

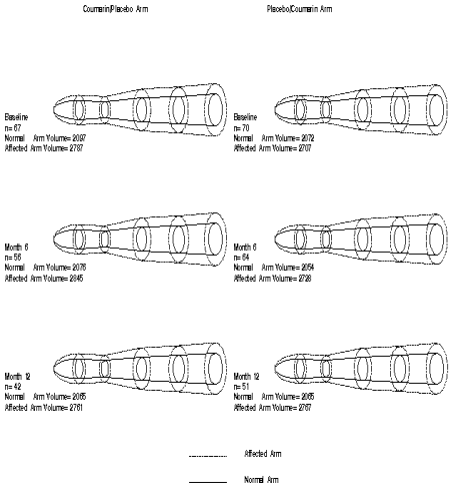


Figure 2: Mean Hotflash Score Reduction for 11 Tx Groups

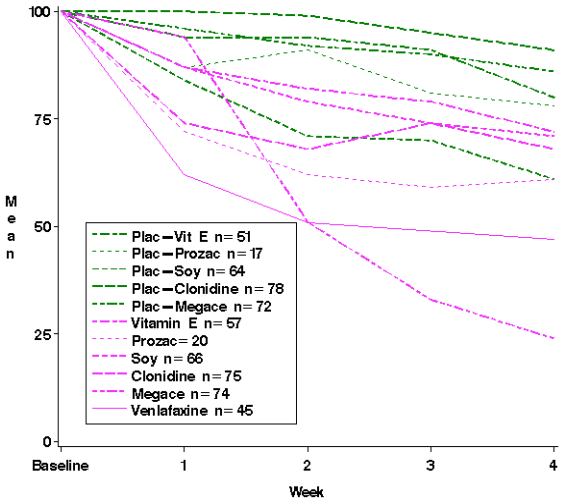


Figure 3: Mean Total Hot Flashes by Day

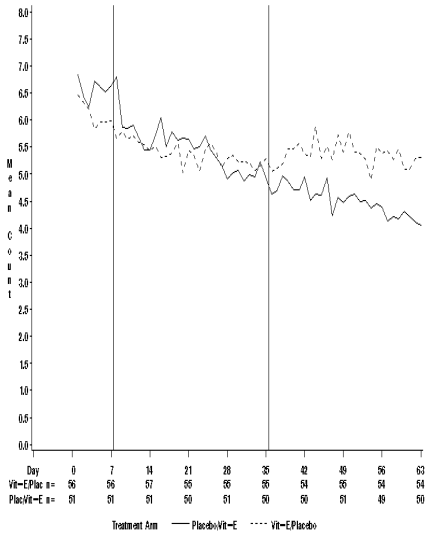


Figure 4: Survival Curves by Physician Estimate of Survival

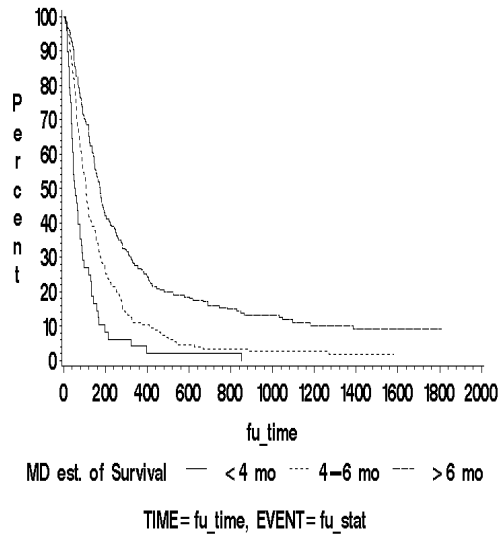


Figure 5: Survival Curves for GBU levels

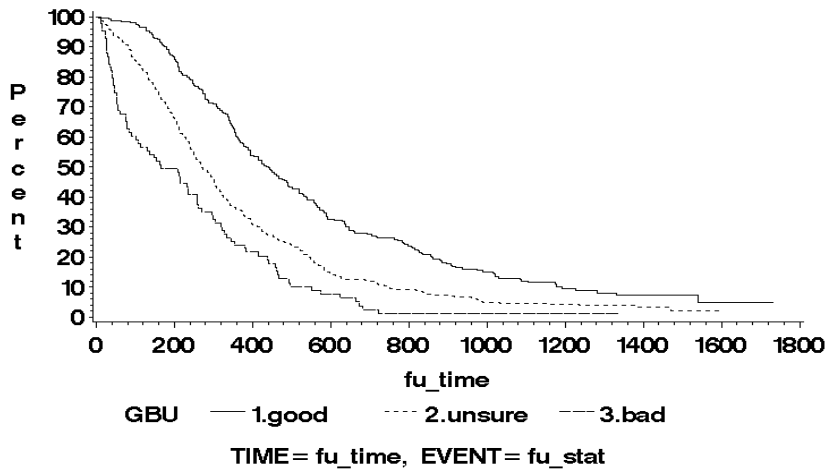


FIGURE 6: CUMULATIVE LISTING OF ADVERSE EVENTS

(as of October 31, 1999)

Patient ID and Initials	Study Phase	Dose at Event (mg/day)	Duration of Drug at Event (days)	Event (description)	Event Grade (NCCTG/NCI Criteria)	Event Date (mo/day/yr)	Event Resolved?	ADR Related to Drug?	Date Off Study (mo/day/yr)	Drop Out Due to ADR?	Serious Adverse Event?
90625 ABC				Breast Tenderness	1	06/16/98	N				N
				Diarrhea	1	06/16/98	N				N
				Tingling	1	06/16/98	N				N
90651 BCD				Stomach Pain	1	03/31/98	N				N
				Hot Flashes	1	02/11/99	N				N
				Impotence	3	02/11/99	N				N
90750 CDE				Genitourinary	1	07/07/98	N				N
90868 DEF				Impotence	1	01/20/98	N				N
				Libido	1	01/20/98	N				N
				PLT NADIR	1	02/05/99	N				N
90963 EFG				Blood in Sperm		06/29/98	N				N
				Breast Tenderness	2	06/29/98	N				N
				Diarrhea	1	06/29/98	N				N
				Dizziness	1	06/29/98	N				N
				Gynecomastia	2	06/29/98	N				N
				Impotence	1	06/29/98	N				N
				Libido	1	06/29/98	N				N
				Loss of Appetite	1	06/29/98	N				N
				Low Back Pain		06/29/98	N				N
90999 FGH				Breast Tenderness	1	07/20/98	N				N
				Gynecomastia	1	07/20/98	N				N
				HGB NADIR	1	12/28/98	N				N
91055 GHI				Headache	1	02/04/98	N		02/04/98		N
				Lower Abd. Pain	1	02/04/98	N		02/04/98		N
				Nausea	1	02/04/98	N		02/04/98		N
				Painful Urination	1	02/04/98	N		02/04/98		N

Prepared for — Protocol Number: 95-92-57
Chemoprevention Branch under Contract N01-CN-65130 — IND Number: 53,704
November 2, 1999
PR=during protocol run in OT=on treatment FU=on follow-up

Figure 7: Mean QOL by Month

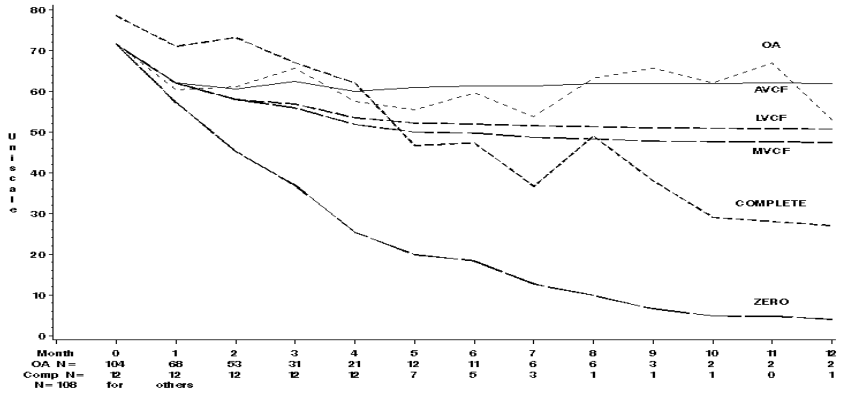


Figure 8: Mean QOL by Month and Arm

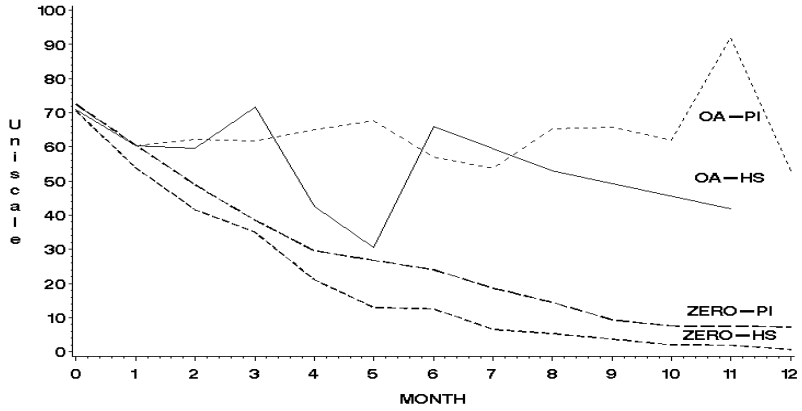


Figure 9: Meg vs. FLU

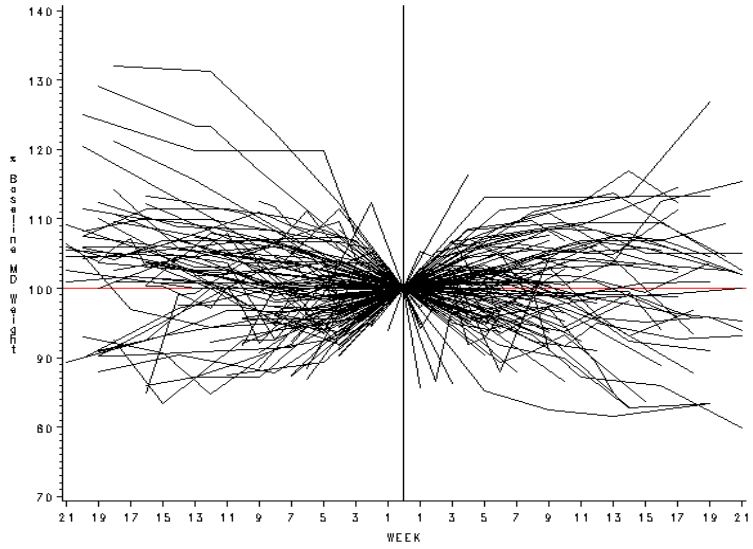
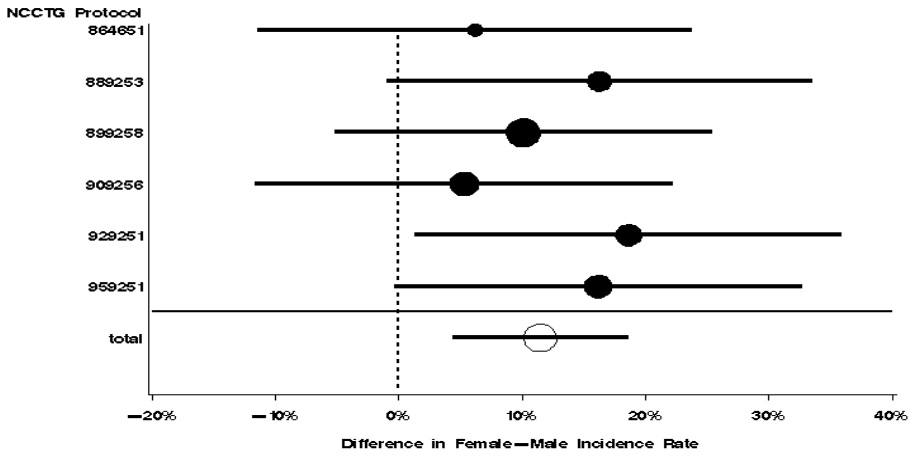


Figure 10: Female Minus Male Difference in Incidence of Any Mouth Soreness with corresponding 95% confidence limits



Note: In the individual studies the size of the circle is proportional to the sample size