TWISTED SURVIVAL: IDENTIFYING SURROGATE ENDPOINTS FOR MORTALITY USING QTWIST AND CONDITIONAL DISEASE FREE SURVIVAL

by

Beth A. Zamboni

BS Statistics, University of Pittsburgh, 1997

MS Biostatistics, Harvard University, 1999

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This dissertation was presented

by

Beth A. Zamboni

It was defended on

April 10, 2015

and approved by

Dissertation Advisor:
Stewart J Anderson, PhD, MA
Professor
Biostatistics
Graduate School of Public Health
University of Pittsburgh

Committee Members:

Ying Ding, PhD
Assistant Professor
Biostatistics
Graduate School of Public Health
University of Pittsburgh

Eleanor Feingold, PhD
Professor
Human Genetics
Graduate School of Public Health
University of Pittsburgh

Yan Lin, PhD
Research Assistant Professor
Biostatistics
Graduate School of Public Health
University of Pittsburgh

Stewart J Anderson, PhD, MA

TWISTED SURVIVAL: IDENTIFYING SURROGATE ENDPOINTS FOR MORTALITY USING QTWIST AND CONDITIONAL DISEASE FREE SURVIVAL

Beth A. Zamboni, PhD

University of Pittsburgh, 2015

ABSTRACT

Traditionally, the standard endpoint in most cancer clinical trials has been overall survival. For many forms of cancer, including colon cancer, the time from diagnosis to the time when this endpoint is reached can take many years. Hence, researchers and patients must wait a considerable amount of time to see if a treatment is effective. We propose an alternative surrogate endpoint which would occur in less time but still be as effective at determining treatment differences. The discovery of such an endpoint would be of Public Health importance to both patients and researchers as it would allow treatments to be tested in a shorter time and subsequently allow patients to have quicker access to a beneficial treatment.

Our approach is to use the methodology of QTWIST and conditional survival estimates, conditioning on disease free survival, to produce a potential surrogate endpoint for overall survival. To do this, we examine whether a surrogate endpoint could be theoretically produced for colon cancer. We analyzed NSABP trials C-03 through C-07 to determine the effect of conditional survival and the choice of conditioning sets in colon cancer. In doing this analysis, we examine the impact of determining the probability of surviving an additional y years given that a patient has already been alive and disease free for x years. QTWIST, quality-adjusted time without symptoms, methodology is then reviewed focusing on the underlying methodology of

using weights, called utility coefficients, and how they could be applied to a partitioning of disease free survival states. Methodology combining conditional survival and the statistical methodology of QTWIST were then performed on six different sets of potential weighting coefficients. Finally, the success of the methodology was evaluated by comparing the Kaplan-Meier treatment difference p-values to the treatment difference p-values for each of the six utility coefficient approaches tested in our methodology. It is our hope that this methodology will produce a viable predictor for overall survival and one that is more predictive than using standard disease free survival estimates.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 OUTLINE OF RESEARCH

Our research is to use the methodology of QTWIST and the conditional survival estimates for overall survival conditional on disease free survival to produce a potential surrogate endpoint for overall survival. In Chapter 2, an introduction to why this research is important to patients, treating physicians and researchers is given. In Chapters 3 and 4, we examine the definition of "surrogate endpoint" presented by other researchers and perform simulations to examine whether these definitions can be applied to colon cancer. Chapter 5 is concerned with looking at the conditional survival estimates for overall survival conditional on disease free survival. The motivating work of Sargent et al. is examined in Section 5.1 and our own previously published work is presented in Section 5.2. In Section 5.3, the original methodology of QTWIST is presented. A description of our methodology, combining our findings of CS and the statistical methodology used in QTWIST, is given in Chapter 6. In Chapter 7, the results from applying our methodology to data obtained from NSABP trials C-03 to C-07 are presented. And finally in Chapter 8, a discussion of our findings are presented.

2.0 INTRODUCTION

It is generally agreed that the most valid way of determining the efficacy of a new treatment relative to the standard treatment for a disease is to perform a randomized trial. In diseases in which the endpoint is reached quickly, this can be a fairly efficient procedure. However, in diseases for which the time from treatment to the time when an endpoint is reached is lengthy, such a trial may take a considerable amount of time. A disease of specific interest to us is cancer, where for moderate risk disease (Stage II and/or Stage III); the time from diagnosis to death from the disease may be many years. Traditionally, the endpoint of interest for virtually all the curable Stage II and III cancer studies has been survival, thus considerable time elapses from the initiation of a trial until there are enough data to accurately assess survival differences. The length of time is somewhat cancer specific. For example, women with Stage II or III breast cancer remain at risk for a recurrence (and subsequent death) for many years after diagnosis, while men and women with stage II or III colon cancer generally die within five years from the cancer or are cured. Since the National Surgical Adjuvant Breast and Bowel Project (NSABP) has performed many adjuvant breast and colon cancer trials in patients with Stage II and III diseases, we have been particularly interested in whether using an endpoint other than survival might both shorten the length of time until the trial can be analyzed, and still be highly likely to accurately predict how the survival endpoints will look when the number of deaths required to assess survival has been reached. There are many reasons why reducing this time would be

useful to patients, treating physicians and researchers. The importance to patients and treating physicians is that the sooner the results of the trial are known, the sooner the information can be used to help them with a treatment. The importance to researchers is that there are numerous new agents to be tested and a limited number of patients (and funds). Thus, if results from a trial can be known earlier, the better treatment can be carried forward into future trials sooner. Our proposal is to try to identify (or refine) a method for doing this with minimal compromise in the strength of the results and the confidence one has in the results.

3.0 REVIEW OF SURROGATE ENDPOINT LITERATURE

The use of surrogate endpoints in clinical trials was a calculated attempt at trying to control the considerable time and expense it takes to complete a clinical trial [3]. This potential benefit to the field of clinical trials led to a series of articles being published in *Statistics in Medicine* in 1989 on the use of surrogate endpoints in clinical trials. In one of these articles, Prentice formally defined a surrogate endpoint in statistical terms. He proposed "that a surrogate for a true endpoint yield a valid test of the null hypothesis of no association between treatment and the true response" [4]. He further stated that this definition required "the surrogate variable to capture any relationship between the treatment and the true endpoint" in essence "requiring the true endpoint at any follow-up time to be independent of treatment". Hence, if one defined T to be the true time to the failure endpoint, W(t) to be the information about the surrogate endpoint for T up to time t, and x to be the covariate identifying the treatment groups being studied then essentially one is requiring

$$\lambda_T\left\{t \mid W(t), X\right\} \equiv \lambda_T\left\{t \mid W(t)\right\} \qquad (1)$$

where the true hazard rate at time t is λ_T [4]. This expression further implies that for each treatment group the relationship between the surrogate and the true endpoint is identical.

Since Prentice's definition was first published much statistical research has been done on the use of surrogate endpoints. Some publications have questioned the overall practice of using surrogate endpoints in both cancer and AIDS clinical trials [5, 6]. Some have suggested the use

of auxiliary endpoints, response variables or covariates that can strengthen true endpoint analyses, as a way to relax Prentice's criteria [7]. While others have proposed alternative study designs and the use of nonparametric tests when analyzing surrogate endpoints [8, 9]. However, the area of surrogate endpoints we choose to explore first was the statistical validation of these endpoints. In the remainder of this chapter, we will use the term biomarker for the potential surrogate endpoint and only use the term surrogate for a biomarker that has met the validation criterion being discussed.

Freedman and Graubard applied Prentice's definition to a binary outcome T (T = 0 or 1) rather than the time to event outcome used by Prentice when initially defining surrogate endpoints [10]. In their work, the biomarker, W, represented an outcome, either discrete or continuous, that was measured before the true endpoint, T, was observed but after the treatment, x, was given. In this new setting, the criterion for assessing a biomarker for surrogacy was

$$P\left(T=1 \mid W, X\right)=P\left(T=1 \mid W\right), \tag{2}$$

which was consistent with Prentice's definition (1).

Using this new expression, they proposed a statistical method for validating the biomarker. Their approach was to use a class of models such as

$$g\left(p\left(T=1 \mid W, X=j\right)\right)=h\left(W\right)+\tau_j \tag{3}$$

where τ_j represents the jth treatment effect. More specifically, they proposed using a logistic transformation for g and the parameters $\mu+\sigma_i$ for h(W) where the parameter μ represents the overall mean and the parameter σ_i represents the effect of the biomarker so that

$$\log\left[\frac{p\left(T=1 \mid W=w_i, X=j\right)}{1-p\left(T=1 \mid W=w_i, X=j\right)}\right]=\mu+\sigma_i+\tau_j \tag{4}$$

Using this model and assuming no interactions, criteria (2) would be satisfied when $\tau_j = 0$. However, they do suggest that before applying this definition, one first test for an interaction between treatment and the biomarker prior to testing $\tau_j = 0$. If there is a significant interaction term then one has strong evidence against (2) and thus there is no need to test the hypothesis $\tau_j = 0$. On the other hand, if there is not a significant interaction term, one should then test the hypothesis. If the treatment effect is significant, then again there is strong evidence against (2). Therefore, one has statistically significant evidence against the use of W as a surrogate endpoint for T.

Freedman and Graubard point out that failure to reject $\tau_j = 0$ is not sufficient to claim the biomarker is a surrogate endpoint. They propose fitting the model once with the biomarker in the model

$$\log\left[\frac{p(T=1 \mid W=w_i, X=j)}{1-p(T=1 \mid W=w_i, X=j)}\right] = \mu + \sigma_i + \tau_{ja} \qquad (5)$$

where τ_{ja} represents the jth treatment effect adjusting for the biomarker

and once without it

$$\log\left[\frac{p(T=1 \mid W=w_i, X=j)}{1-p(T=1 \mid W=w_i, X=j)}\right] = \mu + \tau_j \quad , \qquad (6)$$

then computing the estimate

$$1 - \frac{\hat{\tau}_{1a}}{\hat{\tau}_1} \qquad (7)$$

and using this estimate to measure the proportion of the treatment effect that is explained by the biomarker. Intuitively, the idea is that if adding the biomarker to the model results in the

treatment term being 0, then once the surrogate has been observed, the probability an event will occur is independent of treatment.

In the ideal situation, where criterion (2) is satisfied completely, this new estimate (7) would be 1. They suggested a reasonable ad hoc rule would be to use the biomarker as a surrogate endpoint if the lower limit of a 95% confidence interval for the estimate (7) is greater than 0.5 (or maybe 0.75) and described a method for obtaining the confidence interval using Fieller's Theorem [10].

Lin, Fleming, and De Gruttola applied the Freedman and Graubard approach to the survival setting (the original setting of Prentice's surrogate definition) by identifying the estimate corresponding to (7) and its confidence interval based on a Cox proportional hazards model rather than a logistic model. They suggested fitting two proportional hazards models where W(t) represented a vector of possibly time-varying surrogate covariates. The model with the surrogate endpoint was

$$\lambda(t \mid Z) = \lambda_{20}(t) e^{\beta Z + \gamma' W(t)} \tag{9}$$

And the one without it was

$$\lambda(t \mid Z) = \lambda_{10}(t) e^{\alpha Z} \tag{10}$$

They then defined the proportion of the treatment effect explained by the surrogate endpoint as

$$\hat{p} = 1 - \frac{\hat{\beta}}{\hat{\alpha}} \tag{11}$$

where α, β, and γ are the unknown regression parameters and $\hat{\alpha}$ and $\hat{\beta}$ are their corresponding estimates. They noted that the result is only meaningful if $\hat{\alpha}$ and $\hat{\beta}$ have the

same sign. They used the delta method to show that the random variable $\sqrt{n}\left(\hat{p}-p^{*}\right)$ was asymptotically normal with mean zero and variance

$$\sigma^{2}=\frac{V_{\beta}}{\left(\alpha^{*}\right)^{2}}+\frac{\left(\beta^{*}\right)^{2}V_{\alpha}}{\left(\alpha^{*}\right)^{4}}-2\frac{\left(\beta^{*}\right)V_{\alpha\beta}}{\left(\alpha^{*}\right)^{3}}, \tag{12}$$

where $p^{*}=1-\frac{\beta^{*}}{\alpha^{*}}$ and V_{α}, V_{β}, and $V_{\alpha\beta}$ are the variances and covariance of $n^{\frac{1}{2}}\hat{\alpha}$ and $\mathrm{n}^{\frac{1}{2}}\hat{\beta}$. To obtain estimators of V_{α}, V_{β}, and $V_{\alpha\beta}$, one can use a result from Wei et al. [12] using marginal hazard modeling of multiple events data. A clever method for doing this by constructing artificial bivariate survival data was originally implemented by Lin [13] and [14] and is available in SAS, S-plus and R. We used the program coxph from S-Plus to implement the procedure. For the dependent variable, the survival data was constructed by stringing two copies of the complete survival data. The covariates for the first string of survival data were treatment (to go with α) and two dummy variables of 0 to go with β and γ and for the second string we had a treatment and marker covariate to go with β and γ and a dummy variable of 0 to go with α. It was necessary to use the two strings as strata and patient identifiers as clusters. This allowed us to obtain estimates for α, β, V_{α}, V_{β}, and $V_{\alpha\beta}$. Then one can construct confidence intervals for $\hat{p}$ using standard asymptotic normal theory, i.e. using the interval

$$\hat{p}\pm z_{1-\psi/2}\sqrt{\frac{\hat{\sigma}^{2}}{n}}, \tag{13}$$

Where $\hat{\sigma}$ is obtained by replacing the terms on the RHS of equation 12 with the estimators obtained from coxph. We performed simulations using the same input assumptions as

were used for an example in Lin, Fleming, and De Gruttola and were able to replicate their results (to the degree of accuracy one would expect from simulations).

Lin, Fleming, and De Gruttola also derived the confidence interval applying Fieller's Theorem in this survival setting and obtained

$$1-\left(1-g\right)^{-1}\left[\frac{\hat{\beta}}{\hat{\alpha}}-g\frac{\hat{V}_{\alpha\beta}}{\hat{V}_{\alpha}}\pm\frac{n^{-1/2}z_{1-\psi/2}}{\left|\hat{\alpha}\right|}\sqrt{\hat{V}_{\beta}-2\frac{\hat{\beta}}{\hat{\alpha}}\hat{V}_{\alpha\beta}+\left(\frac{\hat{\beta}}{\hat{\alpha}}\right)^{2}\hat{V}_{\alpha}-g\left(\hat{V}_{\beta}-\frac{\hat{V}_{\alpha\beta}^{2}}{\hat{V}_{\alpha}}\right)}\right] \qquad (14)$$

where $g=z_{1=\psi/2}^{2}n^{-1}\frac{\hat{V}_{\alpha}}{\hat{\alpha}^{2}}$. They pointed out that the confidence interval based on Fieller's theorem was only valid when $g<1$. This implied that the confidence interval based on Fieller's theorem could only be constructed when there was a significant unadjusted treatment effect at the ψ level. The confidence interval based on the normal approximation does not have this constraint, although it is obvious from equation 12 that if the ratio of $\frac{\min(V_{\alpha},V_{\beta})}{\alpha^{2}}$ is large, the variance of $\hat{\sigma}^{2}$ will be large. In fact, if $\frac{V_{\alpha}}{\alpha^{2}}$ is large, the estimate of $\hat{\sigma}^{2}$ will be very unstable. Again, when we used the parameters given in Lin, Fleming, and De Gruttola we were able to replicate their Fieller confidence intervals.

For the simulations of Lin, Fleming, and De Gruttola that we replicated, they generated an equal number of failure times from model (9) for two treatment groups where W was a normal random variable with mean 0 and 2 respectively and unit variance. Furthermore, they let $\lambda_{20}\left(t\right)=1$ and set the restriction that the censoring times be uniformly distributed over 0 and the lower quartile of the simulated failure times.

They then varied γ, $\gamma = \{0.25, 0.5, 1\}$, and n, $n = \{250, 500, 1000\}$, to assess the performance of their statistic. Table 1 shows the results of these simulations. For each of the combinations γ and n, $\hat{\alpha}$, $\hat{\beta}$, their corresponding variances and covariance, $\hat{p}$, $\hat{\sigma}$, and the lower and upper 95% confidence limits for both the normal approximation and Fieller's method were calculated. The mean and variance of these estimates were then calculated based on the 1000 iterations that were performed for each combination of γ and n.

In these simulations the confidence interval based on the normal theory was much tighter than the one obtained using Fieller's Theorem and gave almost identical coverage. More extensive simulations may have identified situations in which the Fieller's Theorem confidence interval performed as well or better than the normal theory confidence interval, but we felt that their findings coupled with some of the previously mentioned negative features of the Fieller's Theorem approach were sufficient to convince us to use the normal theory approach in the remainder of our work..

As before, for criteria (1) to be completely satisfied, p^* would need to be 1. However, as in the binary case, one might decide whether to use the biomarker as a surrogate endpoint if the lower limit of a 95% confidence interval was relatively large, e.g. $\geq .50$. If the lower limit was closer to 0, e.g. < 0.50 one might conclude that there is not sufficient evidence to justify using the biologic marker is a surrogate.

Although the theory developed by Lin, Fleming, and De Gruttola applied to both time dependent and fixed covariates, in their example they only considered a fixed covariate. Since we are interested in using recurrence to predict survival, we chose to simulate some examples that included recurrence as a time-dependent marker, in part to assure that the simulations matched the theory and in part to better understand the concept of surrogacy and how it applies

to the problem of predicting future survival based on recurrence. These simulations are presented in the next chapter.

4.0 SIMULATIONS

Our approach was to initially perform a series of simulations with data conditions similar to what one would see in colon cancer. Since the plan was to better understand the relationship of surrogacy to our desire to predict future results, we began with models that might suit both purposes and then moved to models that delineated the differences between the two goals. For purposes of the following illustrations, we refer to a marker that would allow us to predict a future survival outcome from current marker data as a predictor. To the extent possible, when we generate data, we use parameters that closely match parameter estimates obtained from studying more than 5000 stage II and III colon cancer patients who have received 5 fluorouracil and leucovorin (5FU/LV) on NSABP studies. An assumption that will be used for all of the runs below is that the recurrence distribution for patients receiving the standard regimen is exponential with $\lambda = -(\ln 0.7)/3$. In reality this hazard is only appropriate for the first three years, but for the runs discussed below, we assume that this rate holds throughout the follow-up period. Also in the runs below, we assumed all patients were followed forever without censoring. For simplicity we misuse the term treat and untreated below in the sense that untreated patients will refer to patients receiving 5FU/LV alone (the standard for NSABP trials) and treated patients will refer to those receiving 5FU/LV plus an experimental regimen.

Before presenting any simulations, we note that there is one case for which recurrence would both meet the Prentice criterion for surrogacy (in the work below we call recurrence a

perfect surrogate when this criterion is satisfied) and be an exact predictor. That situation would occur if no one died without recurrence and everyone who recurred would die at a fixed time after recurrence (say two years). Recurrence would then be a perfect surrogate. Furthermore, if time-to-recurrence followed a proportional hazards model, survival would have no hazard for the first two years, but results for recurrence at any time will exactly predict the conditional distribution of survival at t + 2 years given the recurrence data at time t. Thus, a Cox model for time-to-recurrence computed at time t will be identical to the Cox model for survival at time t+2. In this case recurrence is both a surrogate and a perfect predictor.

For our first simulation (Run 1 in Table 2), we generate recurrence data under the assumption that time-to-recurrence follows an exponential distribution for both treated and untreated patients with $\lambda = -(\ln 0.7)/3$ for untreated patients and $\lambda = -(\ln 0.8)/3$ for treated patients. This corresponds to a hazard ratio (treated vs. untreated) of exp (-0.469). We again assume that no one dies without recurrence and we assume that the time from recurrence to death follows an exponential distribution with λ = .25, 1 and 2. Hence it is still true that time to death following recurrence is independent of treatment, so recurrence is still a perfect surrogate. On the other hand, recurrence will not precisely predict survival at a later date, however it appears that the distribution for survival two years later could (theoretically) be obtained and the expected value of the estimated hazard rate for treatment in a survival run at time $t + \Delta$ will be close to the estimated hazard for treatment in a recurrence run at time t for some Δ that may depend on the hazard rate of the distribution of time from recurrence to survival. In this case, recurrence is a surrogate and should be an excellent predictor.

For run 2, although the simulating hazard rates are the same, we have now added a small risk of deaths from other causes (without recurrence) that has a hazard rate of $-(\ln 0.95)/5$ and

is independent of treatment and independent of time-to-recurrence. We assume that once a patient recurs, they no longer are impacted by the competing risk and their survival depends only on the conditional distribution of death following recurrence. In this case, although recurrence is still a perfect surrogate (meets Prentice definition, as well as the p* criterion), it will not be a perfect predictor. In fact, in the recurrence run, the estimated hazard we will obtain for treatment will be larger than that observed for treatment in the subsequent survival analysis (as will the values being estimated). In this example, the survival distribution will not satisfy proportional hazards and the hazard ratio being estimated will be a value between 1 (the hazard for deaths preceding recurrence) and exp(-0.469).

For run 3, the only change was to increase the hazard ratio associated with time to death to $-\left(\ln 0.7\right)/5$. As one can see, the impact was to lower the estimate of treatment effect for survival. This provides an example in which recurrence is again a perfect surrogate, but the treatment parameter from the recurrence run will be a poor predictor of the treatment parameter for the survival run.

In run 4, we simulate data that will illustrate the reverse situation. For this run the hazard ratios for recurrence are again $\lambda = -\left(\ln 0.7\right)/3$ (untreated) and $\lambda = -\left(\ln 0.8\right)/3$ (treated). However this time we assumed that treatment also affected deaths from other causes in the same way by letting the hazard rate for death be $\lambda = -\left(\ln 0.7\right)/5$ for untreated patients and $\lambda = -\left(\ln 0.8\right)/5$ for treated patients. In this case, recurrence will not meet the Prentice definition of surrogacy and p* will not be 1. However, the hazard ratio for treatment for recurrence, exp(-0.469), will still be an excellent predictor of the hazard ratio for survival as the simulations indicate.

These results indicated that if our primary goal is to predict the treatment effect for survival at a future time-point based on current recurrence data, testing for surrogacy is not the correct approach. We will need to focus on the conditional distribution of survival given the current recurrence, survival, and probably additional covariates.

All of the above discussion were based on data having exponential hazard rates and complete follow-up. In the colon cancer setting, we have additional constraints. Some of these constraints make prediction even more useful and plausible, but they also introduce some distribution complexities. Most have these have already been discussed, but a partial listing follows.

1. Time to recurrence does not have a constant hazard. In fact, very few recurrences are observed after three years.
2. The time from recurrence to death is generally fairly short.
3. Other measured time-dependent events occur which have an impact on subsequent survival. For example, second primary cancers, which are collected and recorded, result in some increased risk for death. Many other important events are not collected and/or recorded such as heart attacks, strokes, hip fractures, other events with in sufficient detail to be included etc.
4. We have assumed that our unconditional and conditional time to event distributions are exponential.

For this reason, in the next chapters we will explore conditional survival in colon cancer in more explicit terms. This will help identify the specific problems that need to be addressed and also give us some insight into the way to address them.

5.0 PREVIOUS STATISTICAL METHODOLOGY

5.1 WORK OF DR. DAN SARGENT

In adjuvant colon trials, the usual approach has been to analyze studies approximately five years after the last patient is entered, since this time period will allow one to have nearly complete (colon cancer related) survival data. Previous work done by the NSABP statistical staff has indicated that looking at earlier data that include recurrences and second primary cancers as well as deaths; may allow for an early analysis that will closely mirror that of the survival data that will be observed at five years (from the data of the last patient's entry into the study). The idea is if we had three year data for the endpoints available on all patients we could predict five year overall survival. As we will show below, historical NSABP data has indicated the 3-year disease free survival (DFS) is highly correlated with five year survival results, where the time to event for DFS is the time from surgery (or possibly randomization) to the first of a recurrence, second primary cancer, or death. We will first discuss work performed by Sargent et al. [13] and then present our own published work to support this fact [14].

In 2000, the NSABP opened an adjuvant colon cancer trial, Protocol C-07, which would use disease-free survival obtained after 3 years of follow-up as the primary endpoint. This was the first time a primary endpoint other than survival (with at least 5 years of follow-up) had been utilized. The decision to do this was based on an ad hoc comparison between 3-year DFS and 5-

year survival in previous trials, and the observation that the majority of recurrences occur within the first three years for colon cancer patients, and that the majority of these recurrences will translate into deaths within 3 additional years.

This identification along with conversations with pharmaceuticals, statisticians, and physicians led Dr. Dan Sargent to perform an analysis of 20,898 patients pooled from 18 large randomized phase III colon cancer adjuvant clinical trials whose period of enrollment spanned from 1977-1999 [13]. For each individual protocol, one arm was identified as the control arm whether or not chemotherapy was administered on the arm. A total of 43 different treatment arms were included in the analysis. Furthermore, due to the differing practices in which follow-up was collected; all patients were censored at 8 years from randomization.

Of the 20,898 patients, 19% were 70 or older, 55% were male, 66% were stage III, and 88% received treatment which contained chemotherapy. However, a few of the 20,898 were actually diagnosed with high rectal tumors; these patients were included in the analysis. Eighty-eight percent of the patients who experienced a recurrence during the 8 year follow-up period experienced it within the first 3 years after enrollment. Of these patients, 91% died within 5 years from randomization. The median time from recurrence to death was 12 months. All three of these findings validate what had already been seen in NSABP trials alone.

Using this data, investigators then tested the primary hypothesis that disease free survival (DFS), with 3 years follow-up, was an appropriate endpoint to replace overall survival (OS) with 5 years follow-up in phase III colon adjuvant clinical trials. To test this hypothesis, they first fit a weighted least squares regression line to the DFS and OS pair obtained from each of the 43 study arms. [Figure 1 displays the graph presented in the paper.] The corresponding equation was

5-year OS = -0.02 +1.03 * 3-year DFS

with R^2 and Spearman rank correlation coefficients of 0.85 and 0.88 respectively. However, the intercept and the slope did not significantly differ from 0 and 1 respectively (P>0.20).

Next, they fit a weighted least squares regression line to the within-study hazard ratio comparing the experimental and control arm for DFS and OS. [Figure 2 displays the graph presented in the paper.] The corresponding equation was

OS HR = 0.12 + 0.89*HR DFS

with R^2 and Spearman rank correlation coefficients of 0.90 and 0.94 respectively. However, unlike the DFS vs. OS, here the intercept and the slope do differ significantly from 0 and 1 respectively (P = 0.03 and P < 0.0001 respectively).

Finally, of the 25 within trial (comparing control vs. treatment arms) log-rank tests, 23 yielded the same statistical conclusion using DFS with 3 years of follow-up and OS with 5 years of follow-up. Eighteen agreed no difference and five agreed there was a significant difference. Hence, these findings strengthen the argument for using three year data to predict five year results.

However, there may be a few ways in which Sargent's analyses could have been strengthened. For the Sargent and colleagues's analysis, DFS was defined as the time from randomization to the first event of either recurrent disease or death. Thus their analysis did not account for the different effects on future survival caused by a recurrence versus a second primary versus a death would have on the predictions. Nor did they account for where the recurrence occurred. This type of data, though not available in all institutions, is available through NSABP. We hoped to take advantage of these data to more precisely predict survival

from disease-free survival data in our own work. The results of our work were published in 2010 in the Journal of Clinical Oncology [14] and are presented in the next chapter.

5.2 CONDITIONAL SURVIVAL

Traditionally, survival estimates for patients with colon cancer are reported from the time of diagnosis, Overall Survival (OS). However, these survival projections are not necessarily applicable to patients who have survived a period of time after initial diagnosis. What patients and physicians alike often want to know is "now that one has survived x number of years, what is the chance one will survive another five years?". One approach to answering this question is to compute Conditional Survival (CS) estimates [15]. Several previously published cancer CS data series reveal distinct patterns of CS that vary substantially among diagnoses [16] These studies have explored CS in cancer of the central nervous system (CNS) [17-19], head and neck [20], breast [21-25], lung, gallbladder [29, 30], stomach, colorectum [32-34], prostate [35], ovaries [36], and other sites [37-40]. These studies have found that, in general, CS improves over time most significantly in patients with advanced disease, in whom the survival curve drops off rapidly in the first 1 or 2 years after diagnosis and then stabilizes [15-17, 26, 27, 32, 41].

Conditional Survival (CS) is defined as the probability of survival given that the patient has already survived a certain period of time since diagnosis. CS is derived from the concept of conditional probability in statistics. CS can be calculated from traditional Kaplan-Meier or actuarial life-table survival data. The mathematical definition can be expressed as follows: Let S(t) be the probability a patient survives to time *t*. Conditional survival, CS(y|x), is the

probability of surviving an additional y years, given that the patient has already survived x years, and can be expressed as:

$$CS(y \mid x) = \frac{S(x+y)}{S(x)} = e^{-\int_x^{x+y} h(t)dt} \qquad (15)$$

For example, to compute the 5-year CS for a patient who has already survived 3 years, the absolute survival at 8 years is divided by the absolute survival at 3 years. When a survival curve has a changing hazard rate over time, this will be reflected as a change in CS as more time elapses from diagnosis.

The traditional definition of CS takes into account how long someone has survived but it does not take into account the patient's present disease status relative to recurrence or second primary cancer (of any site). We propose extensions of the concept of CS where one conditions on the set of patients alive and free of recurrence (recurrence-free survival [RFS]) or on the set of patients alive and free of recurrence and second primary cancer (disease-free survival [DFS]). We use notation OS|DFS to generically denote the concept of OS on the set of patients alive and free of disease or OS(y)|DFS(x) for the probability of surviving an additional y years, given that the patient has already been alive and disease-free x years. This expanded definition of CS cannot be directly calculated from survival estimates as CS can, but it can easily be estimated by applying Kaplan-Meier or life-table methods to the patients in the conditioning set. To do this, patients must be followed for recurrence, secondary primary cancer, and death. This definition can further be expanded to other conditioning sets including patients who are alive but have experienced a recurrence or second primary cancer. We refer to these patients as with Disease (D) and use notation OS|D.

In attempt to answer the question on the impact of CS, we performed separate event specific conditional survival analyses on 7004 NSABP patients randomized to receive at least 5FU/LV. Patients in our analysis were subjects in NSABP trials C-03 through C-07, conducted between 1987 and 2002. Only patients from these trials who were enrolled on the FU/LV arms were included in our analysis as the general finding of all of these trials was that FU/LV was more efficacious than comparison chemotherapy regimens. Table 3 displays the numbers of patients from each of the treatments and trials included in the analysis.

Survival, recurrence, and second primary cancer data for all eligible patients with follow-up as well as information on potential prognostic factors such as patient age at diagnosis, gender, race, tumor location, Dukes stage, number of positive nodes, number of nodes resected, and Eastern Cooperative Oncology Group (ECOG) performance status (PS) were included in our analysis.

For all cohorts, we initially calculated 5-year CS, the probability of surviving at least 5 more years as a function of the number of years a patient had already survived since diagnosis, without regard to current disease status. Thus, this analysis did not require the patient to be disease free. It only required the patient to be alive in order to be included in the yearly subset. To account for current disease status, we repeated the analysis for patients alive with disease at 1 to 5 years after diagnosis, OS|D. Finally for a direct comparison to these results and to illustrate the importance of excluding patients with recurrence and second primaries, we calculated 5-year OS|DFS for patients alive and disease-free for up to 5 years after diagnosis. For this analysis, patients were only included in the yearly subset if they had not died, had no recurrence, and had no second primary in the years prior to the given yearly analysis.

All survival times were measured from the time of initial surgery, with the exception of patients in NSABP C-07, for whom date of randomization served as the anchor date. CS estimates were calculated using the Kaplan-Meier method in SAS (SAS Institute, Cary, NC). 95% confidence intervals (CI) were also estimated using SAS.

The impact of prognostic covariates was evaluated by calculating OS|DFS within strata defined by age (≤50, 51-60, 61-70, 71+ years), Dukes stage (B, C), gender, tumor location (left colon, right colon, recto-sigmoid, multiple locations), ECOG PS (0 = normal activity, 1 = symptomatic but ambulatory, 2 = in bed < 50% of time), and number of positive and resected nodes (node negative and ≥12 nodes resected, node negative and <12 nodes resected, 1-3 positive nodes, 4 or more positive nodes).

A total of 6,789 patients were identified in NSABP colon cancer trials C-03 through C-07 as meeting the inclusion criteria for this study (Table 3). Figure 3 shows the results for 5-year CS as a function of years already survived since diagnosis. Error bars depict 95% CI. Without regard to current disease status, the probability of surviving at least 5 more years as a function of the number of years a patient has already survived (5-year CS) increases modestly over the first 5 years, from 77% at diagnosis to 85% at 5 years post-diagnosis. For patients with disease at 1 to 5 years after diagnosis, the 5-year OS|D is very poor, ranging from 7% to 19% (data not graphed). For the set of patients alive and free of disease, the 5-year OS|DFS improves from 77% initially following surgery to 90% after 5 years. One can see that 5-year OS|DFS is consistently greater than the CS estimate. Because the prognosis for patients who have known recurrence or second primary is so much poorer than those who are disease-free, for the remainder of this analysis we restricted our attention to patients who were alive and disease-free for up to 5 years after diagnosis.

Figures 4-7 show the 5-year OS|DFS estimates categorized by the levels of the prognostic covariates age, Dukes stage, number of nodes positive and nodes resected, and PS. No substantial differences in 5-year OS|DFS were observed between males and females or among the tumor location categories (data not shown). For most subsets defined by the levels of prognostic covariates, 5-year OS|DFS estimates increase as time from diagnosis increases. OS|DFS improved with increasing time following diagnosis for all age groups (Fig 4) except for the oldest age group (>70 years), which initially rises from 71% to 78% at 2-3 years, and then falls again back to 70% at 5 years. As expected, Dukes B patients (Fig 5) initially had better survival than Dukes C patients (87% versus 69%), but as time disease-free elapses from diagnosis, OS|DFS for Dukes C improves and approaches the OS|DFS for Dukes B patients (92% versus 88% for Dukes B and C at 5 years, respectively). Patients with 4 or more positive nodes (Fig 6) had markedly worse 5-year survival (57%) at diagnosis compared with other groups (76-89%), but as more time elapses from diagnosis, prognosis improves for all groups and the spread between the groups narrows (86-94%). When grouped by PS (Fig 7), prognosis improves over time for those with a PS of 0 or 1 but not for those with a PS of 2.

In summary the following conclusions were drawn from this analysis:

1. As time progresses from diagnosis, OS|DFS provides more relevant prognostic information for colon cancer survivors than traditional survival prognoses made at the time of diagnosis. In this analysis, we found that improvements in OS|DFS over time were seen for patients with higher stages of disease and for those with more positive nodes. On the other hand, older patients and those with poor PS did not necessarily show an improvement in OS|DFS over time.

2. Differences in survival between strata tend to converge over time. This suggests that certain factors that are well known to have prognostic significance at diagnosis may lose some of their prognostic value as more time elapses from diagnosis.
3. As more time elapses from diagnosis, the number of subjects in each group decreases as patients develop recurrence, second primary cancer, or expire, and some apparent trends observed may actually be due to small patient numbers.
4. CS information is potentially of great interest to patients, their clinicians, and researchers. When patients who are seen in follow-up inquire about their current prognosis, CS can be used to give them a more relevant risk assessment that accounts for time already survived since diagnosis.
5. Changes in prognosis are a result of changing hazard rates over time and can be quantified and easily portrayed using the concept of CS, making the information more accessible to clinicians and more meaningful to patients.
6. We observed little change in 5-year CS after 3 years of DFS in the majority of patient subsets, which is consistent with Sargent's findings.
7. Rates appeared to be exponential for at least the first few years.
8. This analysis further strengthened our belief that if one is looking to propose a surrogate endpoint to survival, the use of conditional survival is a potentially valid possibility.

After the publication of this work, numerous other researchers sited our work in their own studies of conditional survival including studies of renal-cell carcinoma [50, 62], common cancers in Australia [51], colorectal cancer [55, 58], resected colorectal liver metastasis [56], prostate cancer [59], and resected pancreatic adenocarcinoma [63]. Consistently these studies

supported our findings that CS provided more relevant prognostic information for cancer survivors than traditional survival prognoses made at the time of diagnosis. And that controlling for disease status at conditional time periods improves the prediction of subsequent survival.

Additionally, a few researchers sited our work in their research of new statistical methods. Wang et al. developed an interactive tool to make individualized estimates of conditional survival for head and neck cancer patients [53] and rectal cancer patients [54] based on tumor and patient characteristics. Nicolaie et al. researched dynamic predictions in competing risks [57, 60]. Schoop et al. researched measures of prediction error for survival data with longitudinal covariates [52].

Of special note to us was that we could not find research combining our findings of CS and the statistical methodology used in QTWIST. The methodology used in QTWIST will be presented in the next chapter.

5.3 QTWIST

In 1995, Gelber and Cole et al. published a review of the QTWIST (Quality-Adjusted Time Without Symptoms and Toxicity) method [64]. The QTWIST method incorporates quality-of-life considerations into treatment comparisons. The idea was to take overall survival times and partition it into times spent in progressive health states where these particular health states could differ in quality of life. The methodology would then apply threshold utilities, highlighting trade-offs between the different health states, to emphasize treatment comparisons.

In survival analysis, the time to event is from the date of randomization to the date of death. In a QTWIST analysis, the overall survival is broken into multiple time to events

corresponding to changes in quality of life. Weights, called utility coefficients, are then applied to all of the partitioned time to events, called states. Utility coefficients, ranging from zero to one, are then applied to these partitioned time to events where an utility coefficient of zero represents a state as bad as death and an utility coefficient of one represents a state of perfect health. The overall quality-adjusted survival is then found by summing the weighted partitioned time to events.

There are two assumptions made for this method. The first is that "the quality-adjusted time spent in a health state is directly proportional to the actual time spent in the health state, where the proportionality is given by the utility coefficient". The second is that "the utility coefficient for a health state is independent of the time the health state is entered, as well as past and future quality of life".

The application of the QTWIST method is performed in three steps. The first step is to "define quality-of-life oriented survival outcomes that are relevant for the disease setting under study". The assumption is that these outcomes will differ in terms of their time and quality of life and that they will be used to define the progressive health states. The states themselves are assumed to be progressive, however, any of the states can be skipped. Next, utility coefficients are assigned to the health states. And finally, the QTWIST outcome is the sum of the weighted states.

$$QTWIST = u_{State1} * State1 + u_{State2} * State2 + \ldots \qquad (16)$$

To illustrate this first step the following example of outcomes and states are given. The outcomes are listed as TOX (time with toxicity), DFS (disease free survival), and OS (overall survival). TOX is the time when patients are exposed to the toxicities of the therapy. DFS is the time until recurrence or death. OS is the time until death for any reason.

The progressive health states are then TOX (time spent with treatment toxicity), TWIST (time without treatment toxicity and free of disease), REL (time following recurrence). TWIST is calculated as the difference between DFS and TOX (TWIST=DFS-TOX). REL is calculated as the difference between OS and DFS (REL=OS-DFS). All states are subject to right censoring.

The utility coefficients assigned to the states were assigned to reflect the difference between the positive impact that TWIST has on a patient's overall quality of life versus the negative impact that TOX and REL have. For this example, TWIST was given a utility coefficient of 1. The utility coefficients for TOX and REL were unknown so they were represented by u_{TOX} and u_{REL} respectively. However, to illustrate how the use of the utility coefficients could express the difference on the impact on the patient's life. The following example was also provided. If the utility coefficient for TWIST was set at one and the utility coefficient for TOX and REL were set at 0.5 that would represent the situation where "one month spent TOX or REL is equivalent in value to one-half month spent on TWiST".

The QTWIST outcome for this example is then

$$QTWIST = u_{TOX} * TOX + (1) * TWIST + u_{REL} * REL \qquad (17)$$

The second step in the QTWIST method is to "consider each treatment separately and to partition the overall survival time into the defined clinical health states". Then using the Kaplan-Meier product limit method, the areas under the respective curves, TOX, DFS, and OS, are the estimates of the mean health state durations. Thus for their example, the area under the TOX curve is an estimate of the mean duration of TOX. The area between DFS and TOX is an estimate of the mean duration of TWIST. The area between OS and DFS is an estimate of the mean duration of REL. As censoring is almost always present, one cannot estimate the entire survival curve. However, restricted means can be calculated within the follow-up interval of the

study. Bootstrap methods can then be applied to calculation the covariation among these restricted means. Furthermore, partitioned survival plots can be produced to show the transitional survival curves for the multiple outcomes for each separate treatment.

The third step in the QTWIST method is to "compare the treatment regimens in terms of quality-adjust survival (Q-TWiST)". To accomplish this, one combines the estimated restricted mean state durations from Step 2 with the utility coefficients from Step 1. Thus for their example, an array of utility coefficients and the restricted mean estimates for TOX, TWIST, and REL are substituted into equation (17) for each individual treatment. The treatment effect is then estimated by computing the differences in QTWIST between the two treatments. Bootstrap methods can be used to calculate estimates of the standard error. Furthermore, using the Kaplan-Meier large sample theory for restricted means one can conduct statistical inference on the treatment effects.

As this method was designed to incorporate patient's personal preferences of treatment, one can perform a threshold utility analysis displaying treatment comparisons for different values of the utility coefficient. A plot with a threshold line can then be produced to indicate the values of the utility coefficients when the two treatments will have equal QTWIST. A QTWIST gain function can also be calculated by doing the analysis at consistently spaced times up until the end of follow-up. Covariates can also be included in the QTWIST methodology by using proportional hazard models.

After the publication of this work, the QTWIST method was applied in many clinical settings including ovarian carcinoma [65], esophageal carcinoma [66], melanoma [67], breast cancer [68, 69, 73, 83, 88, 92, 95], rectal cancer [70], multiple sclerosis [71, 76], meningitis [72], multiple myeloma [74, 85, 86], leukemia [75, 77], lymphoma [78], cystic fibrosis [79], lung

cancer [80], colorectal cancer [81, 90], epilepsy [82], prostate cancer [84], renal cancer [87], bladder cancer [89], carcinoma of the head and neck [91, 94], and reflux disease [93].

Other researchers have worked on refining the QTWIST method. Zhao et al. derived an estimator for the distribution of quality-adjusted survival time [96]. A year later, Zhao et al. researched how to estimate mean quality adjusted lifetime with censored data [97]. In 2007, Wang and Zhao researched how covariates affect the mean quality of life when the data are subject to right censoring [98]. In 2009, Wang and Zhao did a comparison of the confidence intervals used for mean quality-adjusted lifetime with censored data [99]. Raboud et al. estimated the effect of treatment on quality of life in the presence of missing data due to drop-out and death [100]. Andrei et al. developed generalized linear regression models for the mean of a quality-of-life-adjusted restricted survival time [101]. In 2001, Fine and Gelber investigated a semiparametric bivariate linear regression model for survival and quality-adjusted survival [102]. Bonetti et al. introduced a method-of-moments estimating procedure for categorical quality of life data with nonignorable missingness [103]. Almanassra et al. introduced a new non monotonic estimator for the survival function of quality adjusted lifetime [104].

Again, of special note to us was that we could not find research combining our findings of CS and the statistical methodology used in QTWIST. In the next chapter, we present our research for combining these concepts.

6.0 PROPOSED METHODOLOGY

We propose to use the methodology of QTWIST and the conditional survival estimates for overall survival given disease-free survival (OS|DFS) to produce a potential surrogate endpoint, or at a minimum, a predictor, for overall survival (OS). When using the methodology of QTWIST, we applied the underlying methodology of weighting and not the overarching idea of quality-of-life (QOL). Furthermore, when using OS|DFS we included both recurrence and secondary primaries. To perform this research we followed the steps outlined in the original QTWIST methodology and applied this research to finding predictors in two survival settings, overall survival over the complete follow time (OSC) and overall survival at 5 years OS (OS5). For OSC, we used 5 year disease free survival (DFS5) as a predictor whereas for OS5, we used 3 year disease free survival (DFS3) as a predictor. The following procedure was applied to both settings.

First, we defined the outcomes and states used in QTWIST. The outcomes would be a breakdown of the original QTWIST DFS. For DFS5, we defined the outcomes as being disease free at year 1, year 2, year 3, year 4, and year 5. The states were then: $TWIST_1 = DFS_{Year\ 1}$, $TWIST_2 = DFS_{Year\ 2} - DFS_{Year\ 1}$, $TWIST_3 = DFS_{Year\ 3} - DFS_{Year\ 2}$, $TWIST_4 = DFS_{Year\ 4} - DFS_{Year\ 3}$, $TWIST_5 = DFS_{Year\ 5} - DFS_{Year\ 4}$. For DFS3, we defined the outcomes as being disease free at years 1, 2, and 3. The states were then: $TWIST_1 = DFS_{Year\ 1}$, $TWIST_2 = DFS_{Year\ 2} - DFS_{Year\ 1}$,

$TWIST_3 = DFS_{Year\ 3} - DFS_{Year\ 2}$. In both cases, states were progressive and subject to right censoring.

Next, we assigned utility coefficients to the states. Our initial idea was that the utility coefficients would be assigned to reflect the positive impact that a later TWIST had on a patient's overall survival as compared to an earlier TWIST. Thus, our assumption was that the utility coefficients assigned to each subsequent TWIST state would increase in value. Because we assumed the assignment of utility coefficients was one of the most challenging features to our research, we examined research done by others on the choice of utility coefficients. Revicki et al. [105], Whately-Smith et al. [106], and Tate et al. [107] all made recommendations for measuring utilities but in the traditional QTWIST setting where utility coefficients were based on patient defined utility coefficients or utility coefficients based on health-related quality of life questionnaires and thus did not appear applicable to our research. Revicki et al., however, did caution against "arbitrarily assigning 1.0 to the best possible health state". Thus, we choose to explore a series of utility coefficients where some had the last state's utility coefficient being 1 whereas others did not. We also choose to explore the situation where utility coefficients increase in value for each subsequent TWIST state, as our original plan had stated, and where utility coefficients could increase or decrease as one moves from state to state, as was done in the original QTWIST methods. Finally, we choose to look at utility coefficients based both on the number of events in the state as well as the conditional survival estimates for OS|DFS for the states. In total, six sets of utility coefficients were used to see which would produce a viable predictor.

The first approach was to define the utility coefficients based on the number of TWIST states where the final TWIST states' utility coefficient would be 1. Thus for DFS5, the utility

coefficients were 0.2, 0.4, 0.6, 0.8 and 1. For DFS3, the utility coefficients were 0.33, 0.67, and 1. The second approach was to define the utility coefficients based on the cumulative proportion of disease free events in each state so that the final TWIST states' utility coefficient would again be 1. In the third approach, utility coefficients were calculated based on the proportion of disease free event in each state out of the total number of subjects. Using this idea, the final state's utility coefficient would not be 1.

The final three sets of utility coefficients, approaches 4, 5, and 6, were based on the conditional survival estimates for overall survival conditional on disease free survival, OS|DFS, our original idea for this research. For DFS5, the estimates were based on having a total follow-up time of 5 years. Thus the conditional OSC estimates were based on the probability of surviving an additional *y* years, given that the patient had been disease free for *x* years where x+y=5. For DFS5, the estimates were based on the following five values: CS(5|0), CS(4|1), CS(3|2), CS(2|3), and CS(1|4). Similarly, for DFS3, the estimates were based on having a total follow-up time of 3 years and thus, the conditional OS5 estimates were based on the probability of surviving an additional *y* years, given that the patient had been disease free for *x* years where x+y=3. For DFS3, the estimates were based on the following three values: CS(3|0), CS(2|1), and CS(1|2).

The calculation of all conditional survival estimates were performed in SAS using the Kaplan-Meier method. For approach 4, utility coefficients were set to equal the conditional survival estimates for overall survival conditional on disease free survival, ignoring treatment. In approach 5, we set the utility coefficients to be the absolute value of the absolute difference in the conditional survival estimates for overall survival conditional on disease free survival for the two treatments being compared. The final approach, approach 6, set the utility coefficients to be

the relative difference in the conditional survival estimates for overall survival conditional on disease free survival for the two treatments being compared.

For each of these six approaches, the QTWIST outcome was then calculated. The QTWIST outcome was the sum of the weighted states. For DFS5:

$$QTWIST = u_{TWIST_1} * TWIST_1 + u_{TWIST_2} * TWIST_2 + u_{TWIST_3} * TWIST_3 + u_{TWIST_4} * TWIST_4 + u_{TWIST_5} * TWIST_5 \quad (18)$$

For DFS3:

$$QTWIST = u_{TWIST_1} * TWIST_1 + u_{TWIST_2} * TWIST_2 + u_{TWIST_3} * TWIST_3 \quad (19)$$

Using FORTRAN code, developed by Gelber and Cole, the Kaplan-Meier product limit method was then used to estimate the mean state durations [108]. (We are very thankful to Dr. Gelber and Dr. Cole for granting us permission to use their already developed code.)

For DFS5, the area under the $DFS_{Year\ 1}$ curve is an estimate of the mean of $TWIST_1$. The area between $DFS_{Year\ 2}$ and $DFS_{Year\ 1}$ is an estimate of the mean of $TWIST_2$. The area between $DFS_{Year\ 3}$ and $DFS_{Year\ 2}$ is an estimate of the mean of $TWIST_3$. The area between $DFS_{Year\ 4}$ and $DFS_{Year\ 3}$ is an estimate of the mean of $TWIST_4$. The area between $DFS_{Year\ 5}$ and $DFS_{Year\ 4}$ is an estimate of the mean of $TWIST_5$. For DFS3, estimates for $TWIST_1$, $TWIST_2$, and $TWIST_3$ were similarly calculated.

To make treatment comparisons, the estimates of the mean of the $TWIST_{1\text{-}5\ and}$ $TWIST_{1\text{-}3}$ and the utility coefficients, $u_{TWIST1\text{-}5\ and\ UTWIST1\text{-}3}$, for each of our included treatments were substituted into equation (18) and (19), respectively. Again using the FORTRAN code, the treatment effect was then estimated by computing the differences in QTWIST between the two treatments. P-values for testing the null hypothesis of no treatment effect were calculated based on a Z-test.

As covariates were important in the CS results, we also wanted to look at how covariates affected our research. To do this, we performed our methodology separately on all included trials, C-03 to C-07, as well as separately on all individual covariate profiles. The covariate profiles analyzed were age (≤50, 51-60, 61-70, 71+ years), Dukes stage (B, C), ECOG PS (0 = normal activity, 1 = symptomatic but ambulatory), and number of positive and resected nodes (node negative and ≥12 nodes resected, node negative and <12 nodes resected, 1-3 positive nodes, 4 or more positive nodes).

Finally to evaluate whether the treatment difference in QTWIST was a predictor for treatment difference in overall survival, we compared the Kaplan-Meier -2 log rank p-value for treatment difference in overall survival to the Kaplan-Meier -2 log rank p-value for treatment difference in disease free survival as well as to the QTWIST treatment differences p-value for each of the six utility coefficient approaches. Following Sargent's work, presented in Section 5.1, Spearman rank correlation coefficients were calculated comparing the overall survival p-values to each of the other treatment differences p-values to see which was most highly correlated.

We also wanted to evaluate the methodology based on finding "statistically significant" results. To do this all p-values were classified as being "statistically significant" if the treatment difference p-value was less than 0.05. A comparison of the significance of p-values for overall survival versus disease free survival as well as the six utility coefficient approaches were also performed. In the next chapter, we present our results obtained from applying this methodology to data from NSABP trials C-03 through C-07.

7.0 RESULTS

To determine whether our methodology would produce a viable predictor for overall survival, we tested our methodology on data obtained from NSABP trials C-03 through C-07. For each of these trials, two treatments from each individual trial were included. Thus in addition to the treatments included in our conditional survival analysis, presented in Section 5.2, we also included the treatment regimen MOF for comparison in C-03 and the treatment regimen 5FU+LV+Oxaliplatin for comparison in C-07. To begin the evaluation of our methodology we first began by looking at how effective it was in predicting overall survival over the complete follow time (OSC) based on 5 year disease free survival (DFS5) data. Spearman rank correlation coefficients were calculated based on all treatment p-values, treatment p-values from just the trial level comparisons, and treatment p-values from just the covariate level comparisons. Results from these comparisons are presented in Table 4 and Figure 8.

Examining the predictive value as a whole, none of the methods, including DFS5 appear to be a highly correlated with overall survival as the Spearman rank correlation coefficients varied from 0.49 (DFS5) to 0.62 (TWIST utility 2 - cumulative proportion of disease free events in each state and TWIST utility 3 - proportion of disease free event in each state out of the total number of subjects). However, in all cases, the Spearman correlation coefficient was higher for our methodology then from the traditional DFS5.

A similar pattern was also present when looking at the correlations at the trial level as the correlation for DFS5, 0.51, was smaller than those from our methodology. Again TWIST utility 2 and 3 produced the largest correlation, 0.88, but this time produced a result indicating high association.

When examining the effect at the covariate level, no consistent pattern emerged. For age, Spearman correlation coefficients ranged from 0.36 (TWIST utility 5 - absolute value of the absolute difference in the conditional survival estimates) to 0.63 (TWIST utility 4 - conditional survival estimates ignoring treatment). Dukes stage had a pattern similar to those seen for all p-values and trial level p-values as the correlation for DFS5 was the smallest (0.40) and TWIST utility 2 and 3 were the largest (0.67). For ECOG PS, TWIST utility 5 and 6 resulted in high correlation coefficients (0.81 and 0.85 respectively) and DFS5 had the smallest (0.41). Lastly for number of positive and resected nodes, a different pattern was present. Here TWIST utility 4 and 5 had the smallest values, 0.42, and TWIST utility 1 (number of TWIST states) had the largest value, 0.67.

We also evaluated the methodology based on finding statistically significant results at $\alpha = 0.05$ level. The number of p-values that had agreement and non-agreement are presented in Table 5. Collectively, the number of disagreements for each potential predictor ranged from 8 (TWIST utility 2, 3, and 6) to 9 (DFS5, TWIST utility 1 and 4). At the trial level, there was 1 disagreement in all cases. For age, DFS5 had the fewest disagreements, 2, whereas all others had 3. Dukes stage had the only potential predictor with no disagreements, TWIST utility 5. All other potential predictors had one disagreement. For ECOG PS, most had two disagreements but TWIST utility 5 and TWIST utility 6 each only had one. Lastly, DFS5 had the most

disagreements, 3, when looking at number of positive and resected nodes whereas all the TWIST utility methods had either 1 or 2.

Next we looked to see how effective our methodology was at predicting OS5 based on DFS3 data. Again, Spearman rank correlation coefficients were calculated based on all treatment p-values, treatment p-values from just the trial level comparisons, and treatment p-values from just the covariate level comparisons. Results from these comparisons are presented in Table 6 and Figure 9.

Just as we did for OSC, we began by examining the predictive value as a whole for OS5. In this setting, the Spearman rank correlation coefficients were higher than those calculated for OSC. However, again none of the methods, including DFS3, appeared to be a highly correlated with OS5 as the Spearman rank correlation coefficients varied from 0.56 (TWIST utility 2) to 0.67 (3 year disease free survival and TWIST utility 4). This pattern was opposite of what we saw for OSC.

At the trial level, DFS3 as well as four of the TWIST utility methods (1, 2, 3, and 4) all produced large correlation values with TWIST utility 2 and 3 producing the largest (0.94). The result that TWIST utility 2 produced the largest correlation was also seen for OSC.

When examining the effect at the covariate level for OS5, again no consistent pattern emerged. For age, Spearman correlation coefficients ranged from 0.49 (TWIST utility 2 and 3) to 0.73 (TWIST utility 6 - relative difference in the conditional survival estimates). Dukes stage had correlations varying from 0.68 (TWIST utility 2, 3, and 6) to 0.80 (DFS3). For ECOG PS, correlations ranged from 0.69 (TWIST utility 2 and 3) to 0.87 (TWIST utility 6). Lastly for number of positive and resected nodes, another pattern was present. Here TWIST utility 2 and 3 had the smallest values, 0.43, and TWIST utility 5 had the largest value, 0.67.

In mirroring the evaluation we did for OSC, next we evaluated our methodology for OS5 based on finding statistically significant results at $\alpha = 0.05$ level. The number of p-values that had agreement and non-agreement are presented in Table 7. Collectively, the number of disagreements for each potential predictor ranged from 6 (TWIST utility 5 and 6) to 11 (TWIST utility 1). At the trial level, TWIST utility 5 and 6 had no disagreements whereas all others had 1. For age, 5 year DFS had the most disagreements, 3, whereas all others had 1 or 2. Dukes stage had 2 disagreements for all cases. TWIST utility 4, 5, and 6 had no disagreements for ECOG PS while the remainder had 1 or 2. Lastly, the covariate number of positive and resected nodes had between 3 and 4 disagreements for all potential predictors.

8.0 DISCUSSION

Because one of the main goals of our research was to identify a methodological approach that would produce a viable predictor of overall survival, we were disappointed by the range of Spearman rank correlation coefficients comparing overall survival treatment p-values to disease free treatment p-values. However, our methodology did produce correlation coefficients that were generally higher than seen using a standard DFS approach. Our methodology also produced a greater amount of agreement in statistically significant findings then using a standard DFS surrogate approach.

In particular, one utility coefficient approach appeared to produce that most viable predictors: TWIST utility 5 - absolute value of the absolute difference in the conditional survival estimates. For TWIST utility 5, the number of disagreements in statistical significance were seven (11.1%) for OSC to DFS5 and six (9.5%) for OS5 to DFS3.

For TWIST utility 5, six of the seven disagreements for OSC compared to DFS5 resulted in OSC finding a non-statistically significant result whereas the DFS5 p-value was less than 0.05. A further examination of the OSC Kaplan Meier plots for five of these cases showed similar patterns in their survival curves. For these five, the curves started together, split apart, and then came back together. The exception to this pattern was the trial level case with the smallest amount of follow-up time.

When using TWIST utility 5 to compare results for OS5 to DFS3, a total of six disagreements in statistically significance occurred with 3 occurring in each direction. In two of the three where OS5 was not statistically significant but DFS3 was, we see a similar pattern to what we saw for OSC. A different pattern is visible when looking at two of the three cases where OS5 was statistically significant but DFS3 was not. In these two, the separation in the survival curves does not occur until after three years of follow-up.

In addition to TWIST utility 5 producing the fewest statistically significant disagreements, this approach also produced relatively high Spearman correlation coefficients when comparing OS5 treatment p-values to 3 year disease free treatment p-values (0.64 to 0.79) with the one exception being the Spearman correlation coefficients for trial level comparisons (0.38). However, all trial level comparisons did result in agreement in whether the treatment was statistically significant or not.

In conclusion, it does appear that our methodology, combining the methodology of QTWIST and weighting coefficients determined by conditional survival, does produce a potential predictor of overall survival and one that may be more predictive than standard DFS methodology.

APPENDIX A: FIGURES AND TABLES

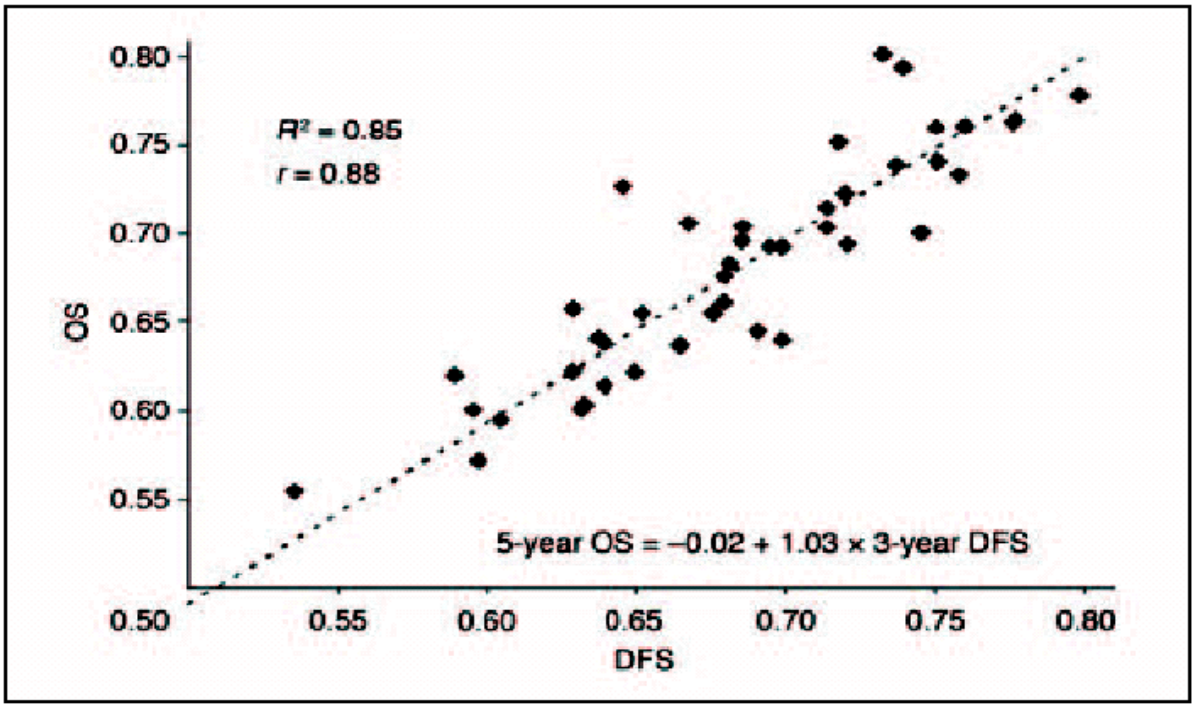

Figure 1. Three-year disease-free survival (DFS) vs 5-year overall survival (OS) by study arm R

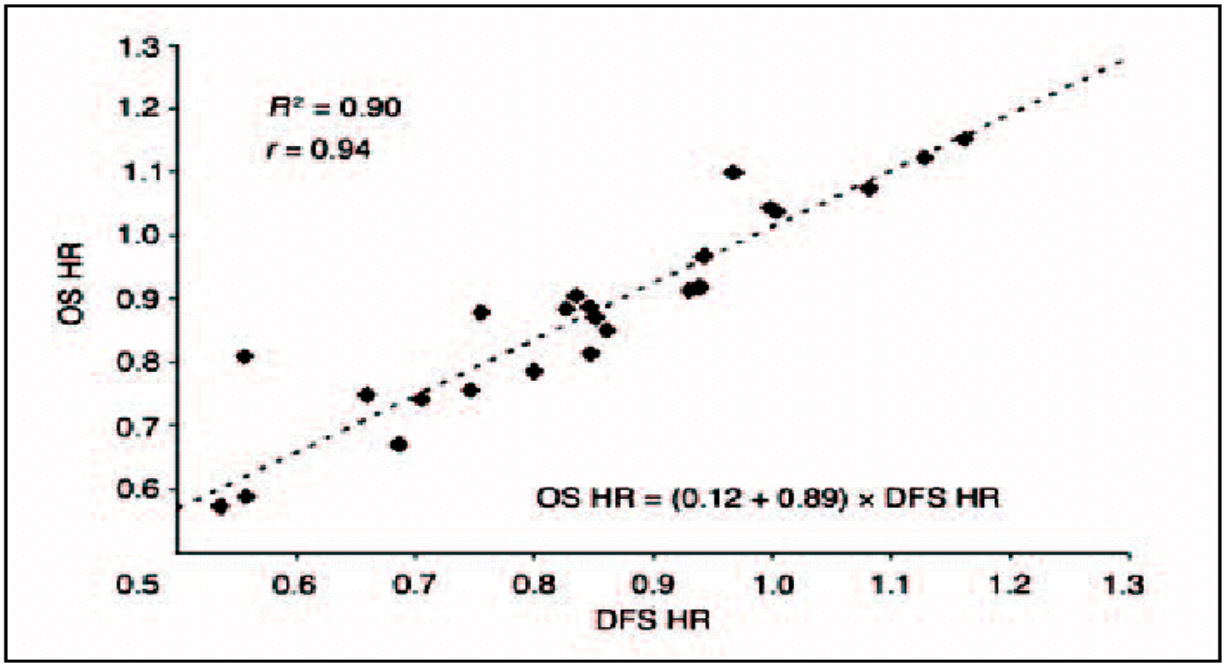

Figure 2. Disease-free survival (DFS) versus overall survival (OS) hazard ratios (HR) by trial

Table 1. Summary statistics for Lin, Fleming and De Gruttola simulation studies (1000 simulation samples)

	$\gamma = 0.25$			$\gamma = 0.50$			$\gamma = 1.00$		
N	250	500	1000	250	500	1000	250	500	1000
$Mean(\hat{\alpha})$	1.56	1.52	1.50	2.15	2.01	1.97	3.47	2.91	2.78
$SE(\hat{\alpha})$	0.47	0.31	0.22	1.10	0.37	0.26	2.47	1.03	0.35
$Mean(\hat{\beta})$	1.07	1.03	1.01	1.18	1.05	1.02	1.69	1.16	1.04
$SE(\hat{\beta})$	0.58	0.38	0.28	1.17	0.43	0.31	2.48	1.05	0.39
$Corr(\hat{\alpha},\hat{\beta})$	0.80	0.79	0.81	0.96	0.84	0.85	0.99	0.98	0.91
$Mean(\hat{p})$	0.35	0.33	0.33	0.50	0.49	0.49	0.62	0.63	0.64
$SE(\hat{p})$	0.31	0.17	0.13	0.24	0.15	0.11	0.22	0.14	0.10
$Mean(n^{-1/2}\hat{\sigma})$	0.28	0.18	0.13	0.22	0.15	0.11	0.19	0.13	0.09
Mean width of 95 per cent CI									
$\delta - method$	1.11	0.72	0.50	0.87	0.60	0.42	0.73	0.52	0.37
Fieller method	1.69	0.84	0.52	1.30	0.66	0.44	0.90	0.57	0.38
Coverage of 95 per cent CI									
$\delta - method$	0.96	0.96	0.96	0.94	0.95	0.95	0.90	0.94	0.94
Fieller method	0.94	0.96	0.96	0.94	0.96	0.95	0.91	0.96	0.95

Table 2. Results for simulated cancer examples

		Without Interaction				With Interaction			
	λ^{RS}	α	β	$1-\frac{\beta}{\alpha}$	γ	α	β	γ	$\beta * \gamma$
Run 1: $\lambda_o^S =$ $\lambda_T^S =$ $\lambda_o^R = -(\ln 0.7)/3$ $\lambda_T^R = -(\ln 0.8)/3$	0.25	**-.440**	**.001**	**1.002**	8.58	-.440	.116	8.67	-.115
	1	**-.467**	**.002**	**1.004**	11.87	-.467	.090	13.36	-.088
	2	**-.470**	**-.001**	**0.998**	21.00	-.470	.455	23.80	-.451
Run 2: $\lambda_o^S = -(\ln 0.95)/5$ $\lambda_T^S = -(\ln 0.95)/5$ $\lambda_o^R = -(\ln 0.7)/3$ $\lambda_T^R = -(\ln 0.8)/3$	0.25	**-.378**	**-.002**	**0.995**	3.20	-.378	.004	3.20	-.006
	1	**-.417**	**-.001**	**0.998**	4.56	-.417	.019	4.57	-.027
Run 3: $\lambda_o^S = -(\ln 0.7)/5$ $\lambda_T^S = -(\ln 0.7)/5$ $\lambda_o^R = -(\ln 0.7)/3$ $\lambda_T^R = -(\ln 0.8)/3$	0.25	**-.168**	**.001**	**1.006**	1.26	-.168	.000	1.25	.001
	1	**-.242**	**.002**	**1.008**	2.66	-.242	.004	2.67	-.004
	3	**-.264**	**-.003**	**0.989**	3.77	-.265	-.005	3.77	-.007
Run 4: $\lambda_o^S = -(\ln 0.7)/5$ $\lambda_T^S = -(\ln 0.8)/5$ $\lambda_o^R = -(\ln 0.7)/3$ $\lambda_T^R = -(\ln 0.8)/3$	0.25	**-.405**	**-.175**	**0.568**	1.50	-.405	-.468	1.25	.468
	1	**-.463**	**-.180**	**0.611**	2.91	-.463	-.470	2.67	.468
	3	**-.466**	**-.185**	**0.603**	3.55	-.467	-.394	3.64	.381

Table 3. Description of the number of patients included in colon cancer analysis

Study	Arm	# Randomized	# Ineligible	# No Follow-up	# Included
C-03	LV + 5FU	539	19	2	518
C-04	LV + 5FU	719	26	2	691
C-04	LV+5FU+LEV	717	20	2	695
C-05	LV + 5FU	1088	18	1	1069
C-05	LV+5FU+IFN	1088	21	8	1059
C-06	LV + 5FU	803	27	6	772
C-06	LV + UFT	805	24	1	783
C-07	LV + FU	1245	23	15	1207
Total		7004	178	37	6789

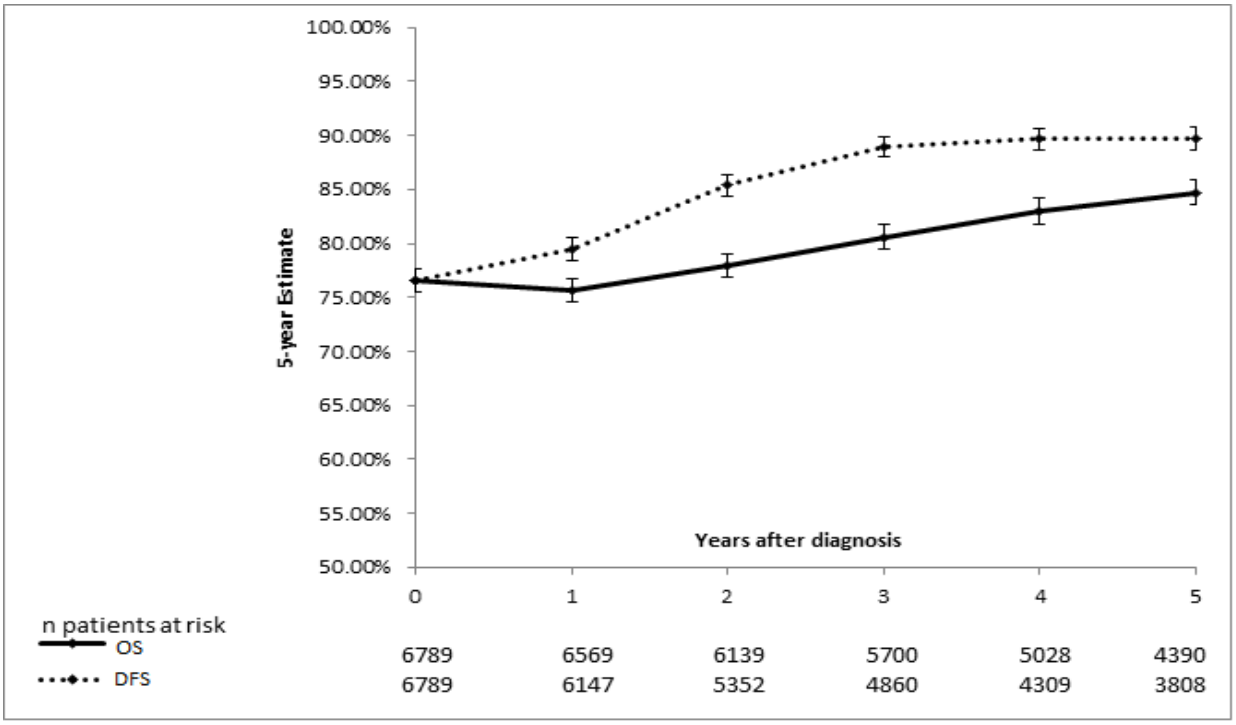

Figure 3. Five year CS as a function of years already survived since diagnosis

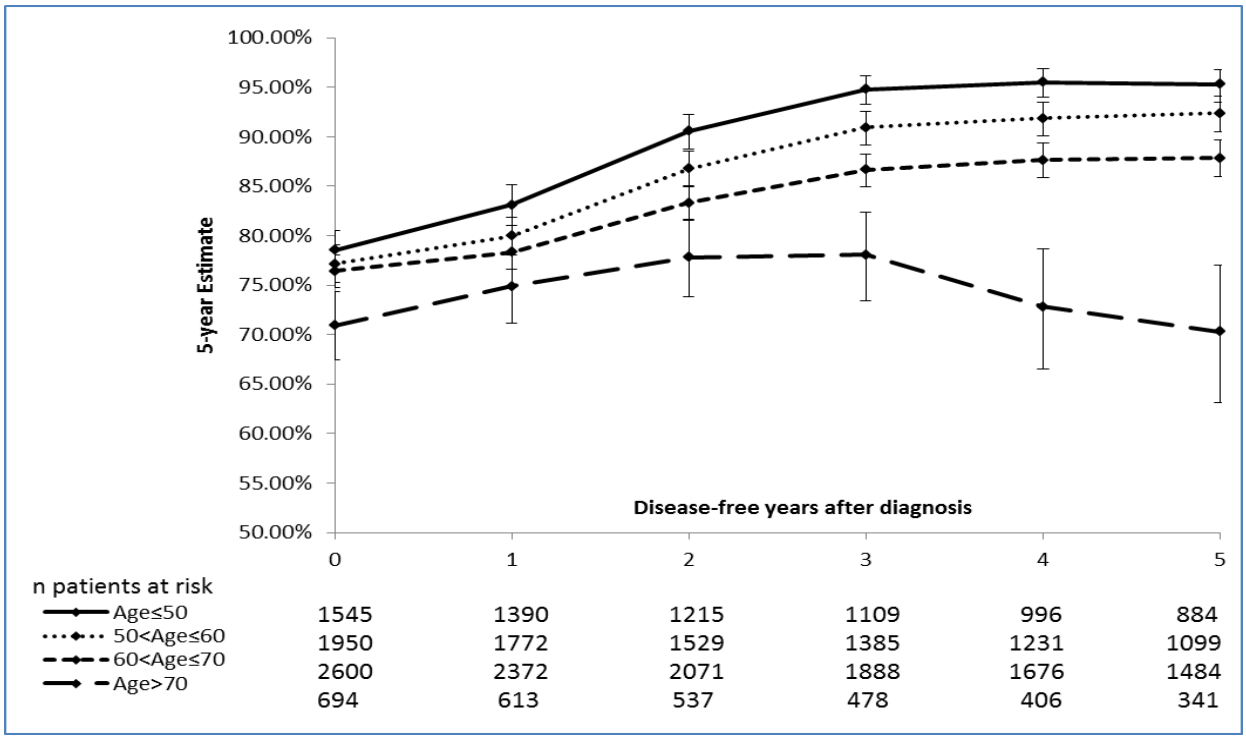

Figure 4. Five year OS|DFS estimates categorized by age

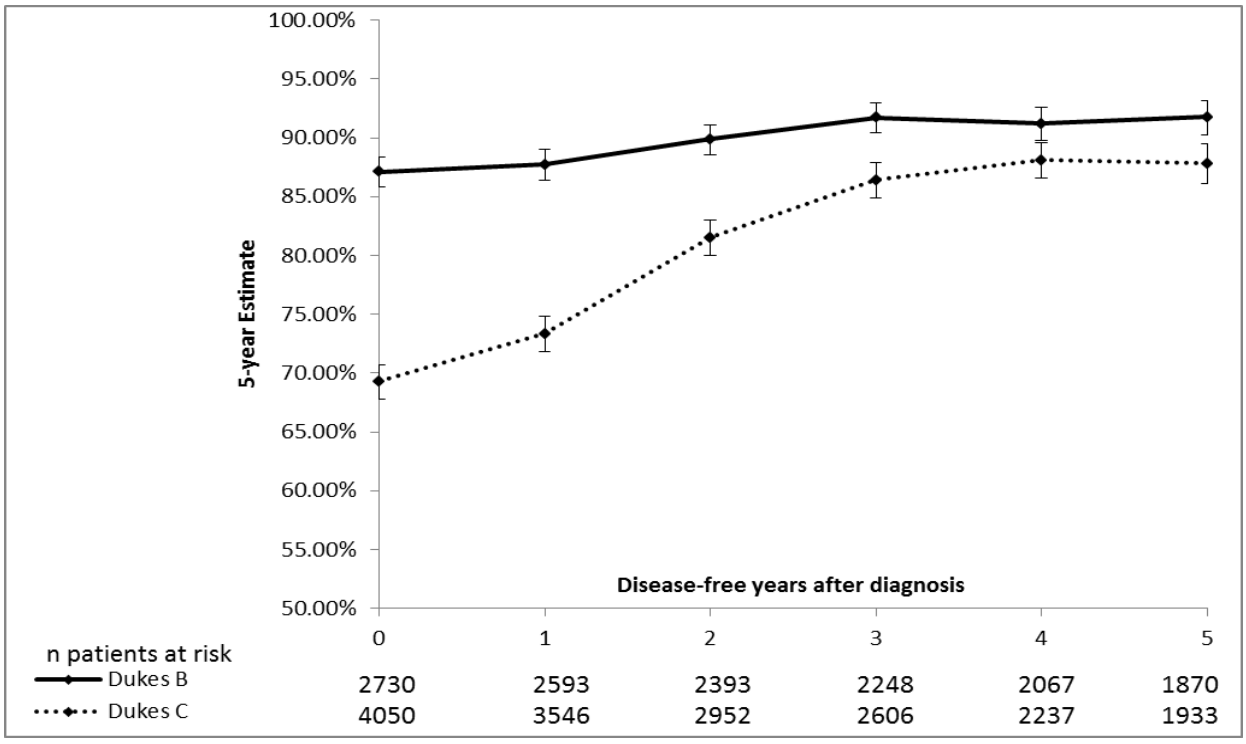

Figure 5. Five year OS|DFS estimates categorized by Dukes stage

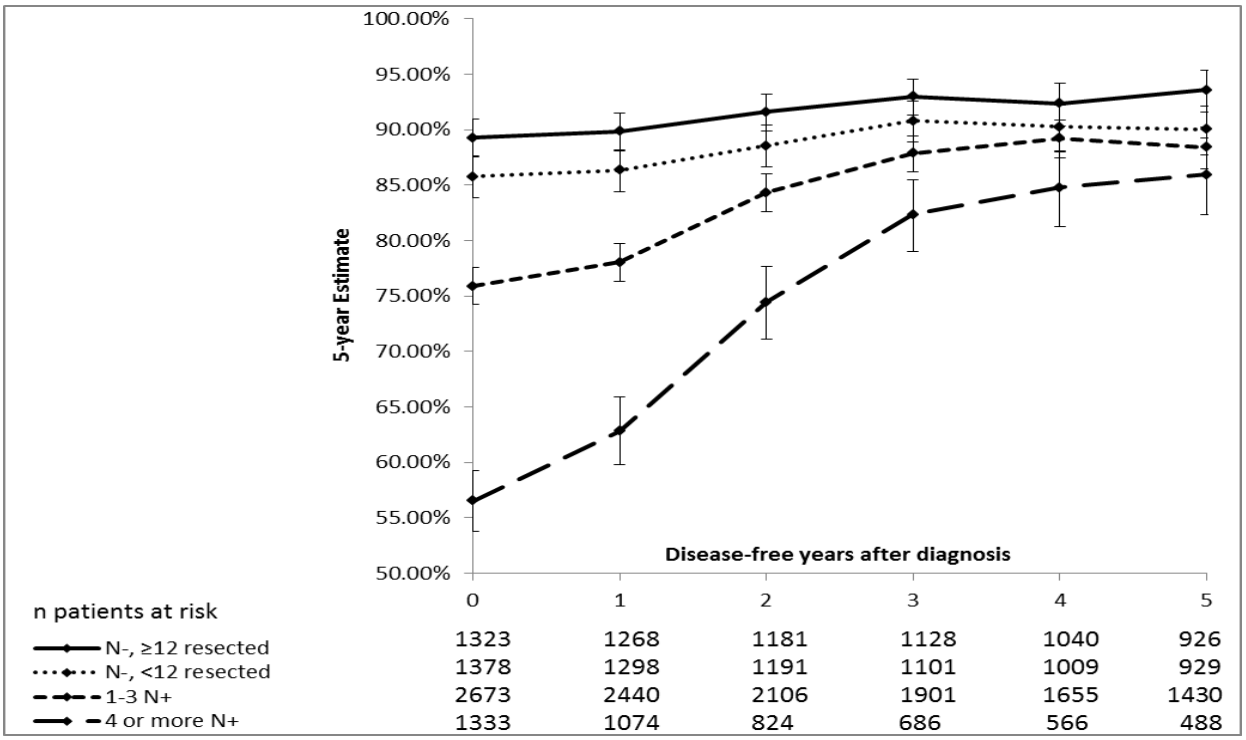

Figure 6. Five year OS|DFS estimates categorized by positive and resected nodes

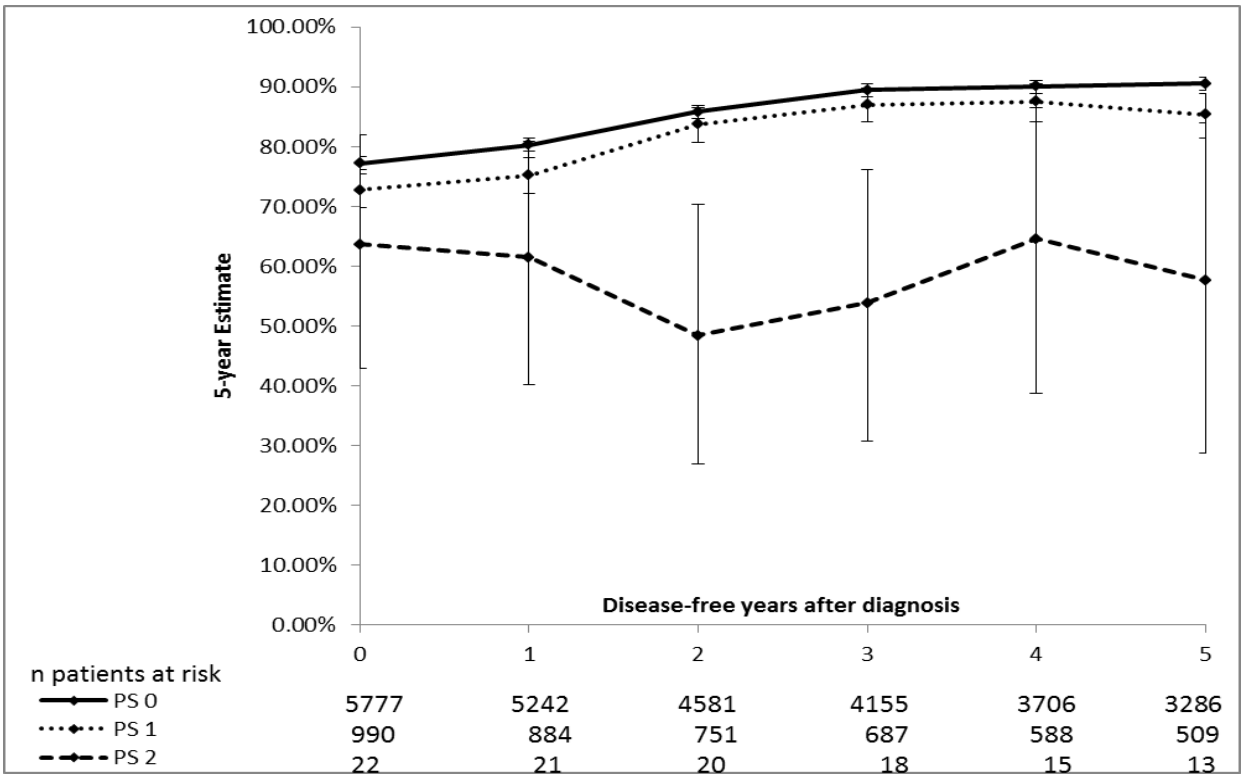

Figure 7. Five year OS|DFS estimates categorized by ECOG PS

Table 4. **Spearman rank correlation coefficients comparing overall survival (OSC) p-values to 5 year disease free survival** (DFS5)

	N	DFS5	TWIST Utility 1	TWIST Utility 2	TWIST Utility 3	TWIST Utility 4	TWIST Utility 5	TWIST Utility 6
All	63	0.49	0.61	0.62	0.62	0.50	0.49	0.50
Trial level	5	0.51	0.84	0.88	0.88	0.64	0.63	0.57
Age	18	0.62	0.55	0.56	0.55	0.63	0.36	0.42
Dukes stage	10	0.40	0.63	0.67	0.67	0.55	0.43	0.53
ECOG PS	10	0.41	0.58	0.63	0.63	0.48	0.81	0.85
# of positive and resected nodes	20	0.50	0.67	0.64	0.64	0.42	0.42	0.38

Table 5. Number of p-values with agreement (+) and non-agreement (-), α=0.05 level, for overall survival (OSC) versus 5 year disease free survival (DFS5)

	DFS5		TWIST Utility 1		TWIST Utility 2		TWIST Utility 3		TWIST Utility 4		TWIST Utility 5		TWIST Utility 6	
	+	-	+	-	+	-	+	-	+	-	+	-	+	-
All	54	9	54	9	55	8	55	8	54	9	56	7	55	8
Trial level	4	1	4	1	4	1	4	1	4	1	4	1	4	1
Age	16	2	15	3	15	3	15	3	15	3	14	4	15	3
Dukes stage	9	1	9	1	9	1	9	1	9	1	10	0	9	1
ECOG PS	8	2	8	2	8	2	8	2	8	2	9	1	9	1
# of positive and resected nodes	17	3	18	2	19	1	19	1	18	2	19	1	18	2

Table 6. Spearman rank correlation coefficients comparing 5 year overall survival (OS5) p-values to 3 year disease free survival (DFS3)

	N	DFS3	TWIST Utility 1	TWIST Utility 2	TWIST Utility 3	TWIST Utility 4	TWIST Utility 5	TWIST Utility 6
All	63	0.67	0.60	0.56	0.56	0.67	0.64	0.64
Trial level	5	0.88	0.93	0.94	0.94	0.87	0.38	0.36
Age	18	0.63	0.53	0.49	0.49	0.68	0.70	0.73
Dukes stage	10	0.80	0.70	0.68	0.68	0.79	0.69	0.68
ECOG PS	10	0.78	0.72	0.69	0.69	0.72	0.79	0.87
# of positive and resected nodes	20	0.59	0.50	0.43	0.43	0.62	0.67	0.64

Table 7. Number of p-values with agreement (+) and non-agreement (-), α=0.05 level, for 5 year overall survival (OS5) versus 3 year disease free survival (DFS3)

	DFS3		TWIST Utility 1		TWIST Utility 2		TWIST Utility 3		TWIST Utility 4		TWIST Utility 5		TWIST Utility 6	
	+	-	+	-	+	-	+	-	+	-	+	-	+	-
All	53	10	52	11	53	10	53	10	55	8	57	6	57	6
Trial level	4	1	4	1	4	1	4	1	4	1	5	0	5	0
Age	15	3	16	2	17	1	17	1	16	2	17	1	17	1
Dukes stage	8	2	8	2	8	2	8	2	8	2	8	2	8	2
ECOG PS	9	1	8	2	8	2	8	2	10	0	10	0	10	0
# of positive and resected nodes	17	3	16	4	16	4	16	4	17	3	17	3	17	3

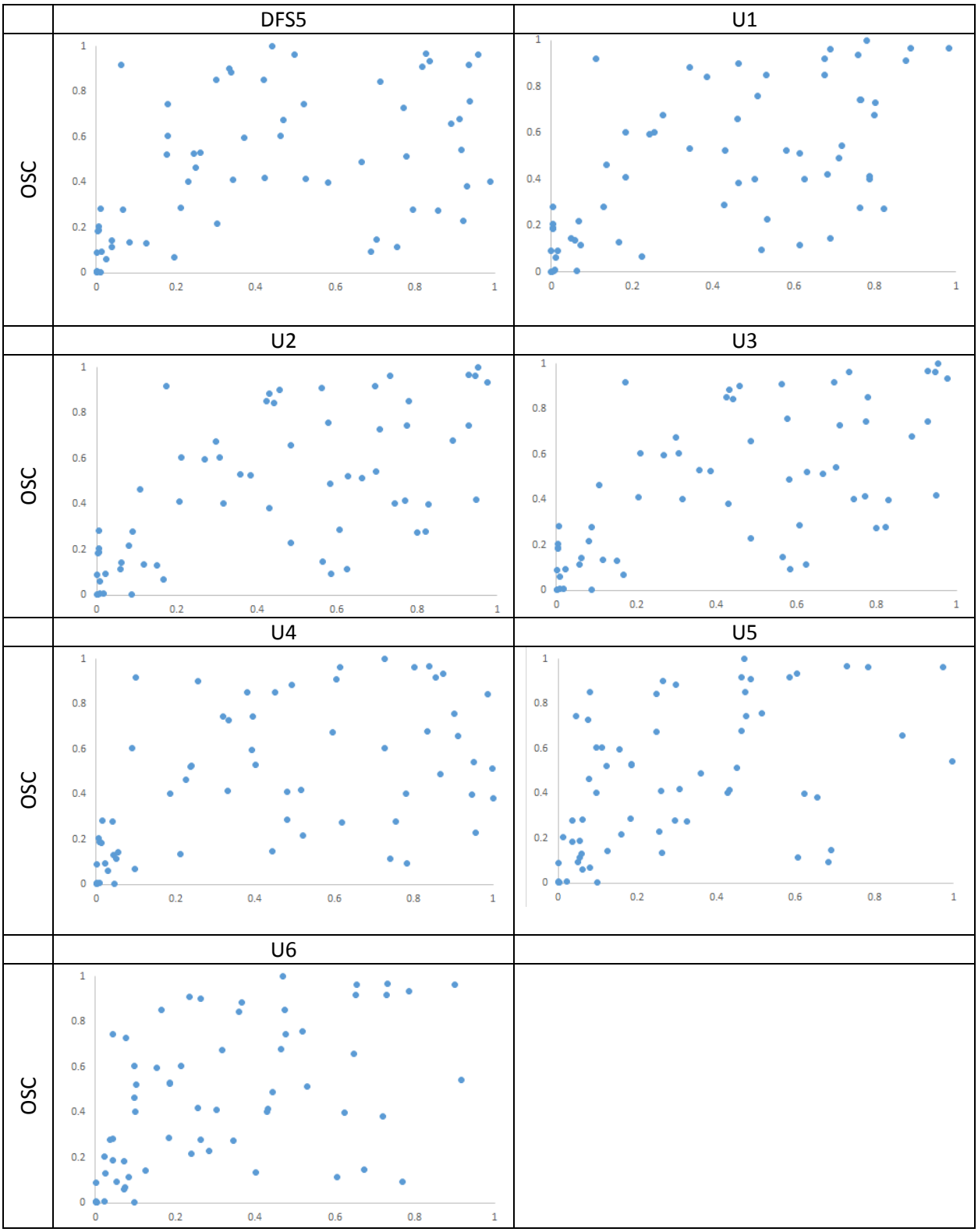

Figure 8. Comparison of treatment difference p-values for OSC to DFS5

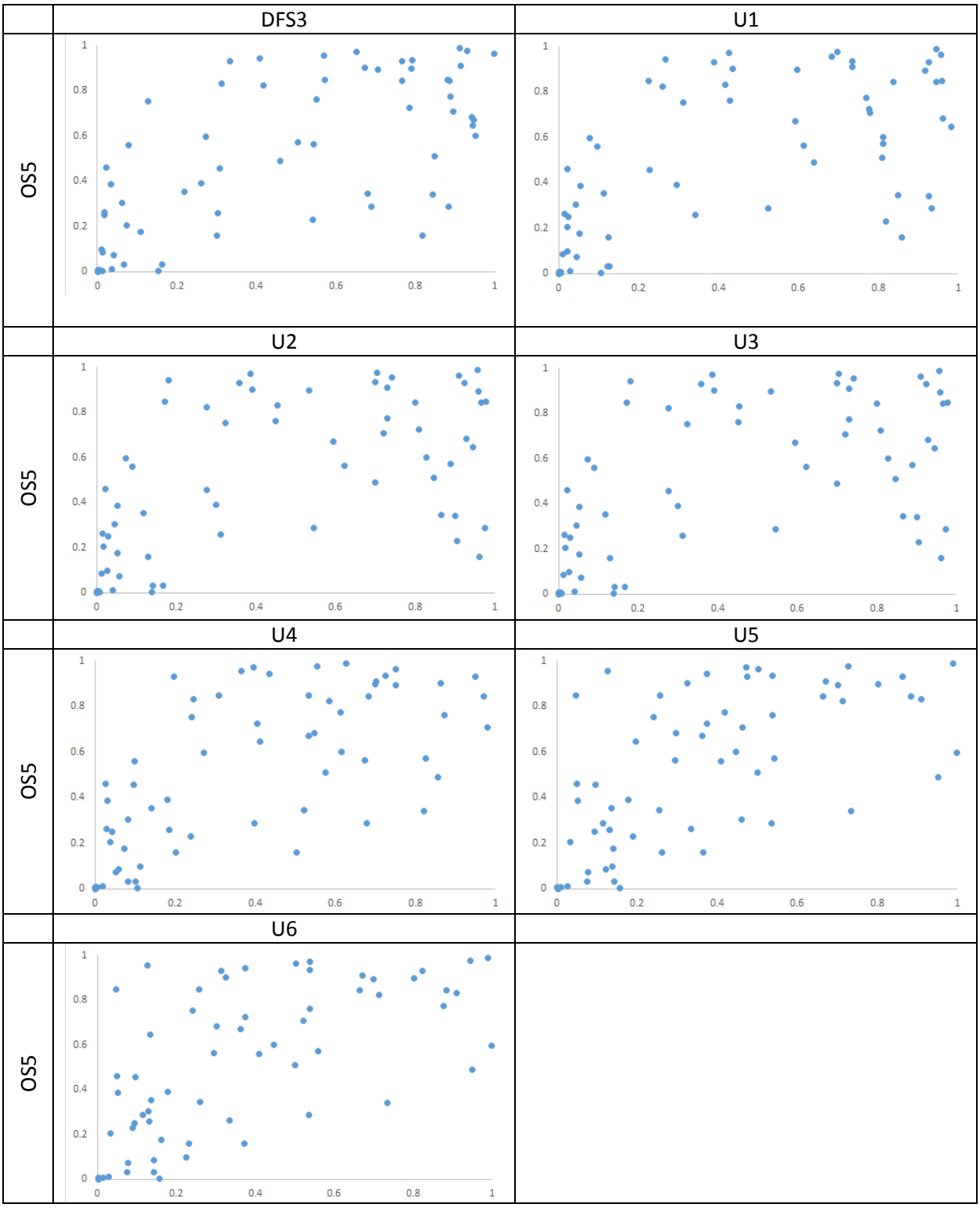

Figure 9. Comparison of treatment difference p-values for OS5 to DFS3

APPENDIX B: STATISTICAL CODE

B.1 LIN, FLEMING AND DE GRUTTOLA SIMULATION CODE

```
niter_1000

n0_250
n1_250
n_n0+n1

id_rep(1:n,1)
id1_c(id,id)

u0_0
u1_2

beta_1
delta_0.50
lambda20_1

z_rep(1.96,niter)
nf_rep(n,niter)

r_c(rep(0,n0),rep(1,n1))
r1_c(r,rep(0,n))
r2_c(rep(0,n),r)

etype_c(rep(1,n),rep(2,n))

modelalpha_c(rep(0,niter))
modelvaralpha_c(rep(0,niter))
modelbeta_c(rep(0,niter))
modelvarbeta_c(rep(0,niter))
modelcov_c(rep(0,niter))

psim_c(rep(0,niter))
sigma2_c(rep(0,niter))
nsigma_c(rep(0,niter))
sigma2check_c(rep(0,niter))
sigma2new_c(rep(0,niter))
```

```
nsigmanew_c(rep(0,niter))

lowernci_c(rep(0,niter))
uppernci_c(rep(0,niter))
widthnci_c(rep(0,niter))
coveragen_c(rep(0,niter))

g_c(rep(0,niter))
q_c(rep(0,niter))
lowerfci_c(rep(0,niter))
upperfci_c(rep(0,niter))
widthfci_c(rep(0,niter))

for (i in 1:niter)
{
w_c(rnorm(n0,u0,1),rnorm(n1,u1,1))
w1_c(rep(0,n),w)

lambda_lambda20*exp((beta*r)+(delta*w))
nlambda_rep(1,n)
timevar_c(rexp(n,lambda))
timevar1_c(timevar,timevar)
censor_runif(n,0,quantile(timevar,0.25))
censor1_c(censor,censor)

status1_c(rep(0,n),rep(0,n))
status1[timevar<censor]_1

model_coxph(Surv(timevar1,status1)~(r1 + r2 +
w1)*strata(etype)+cluster(id1))
modelalpha[i]_model$coef[1]
modelvaralpha[i]_model$var[1]
modelbeta[i]_model$coef[2]
modelvarbeta[i]_model$var[8]
modelcov[i]_model$var[7]

psim[i]_1-(modelbeta[i]/modelalpha[i])
sigma2[i]_((modelvarbeta[i]/modelalpha[i]^2)+(((modelbeta[i]^2)*modelv
aralpha[i])/modelalpha[i]^4)
      -(2*modelbeta[i]*modelcov[i]/modelalpha[i]^3))
nsigma[i]_sqrt(sigma2[i])

lowernci[i]_psim[i]-z[i]*nsigma[i]
uppernci[i]_psim[i]+z[i]*nsigma[i]
widthnci[i]_uppernci[i]-lowernci[i]
if (lowernci[i]<0.33 && uppernci[i]>0.33) coveragen[i]_1

g[i]_z[i]^2*(modelvaralpha[i]/modelalpha[i]^2)
q[i]_modelbeta[i]/modelalpha[i]
upperfci[i]_1-(1/(1-g[i]))*(q[i]-
(g[i]*(modelcov[i]/modelvaralpha[i]))-(z[i]/abs(modelalpha[i]))
```

```
     *(modelvarbeta[i]-
(2*q[i]*modelcov[i])+((q[i]^2)*modelvaralpha[i])-
(g[i]*(modelvarbeta[i]-
     (modelcov[i]^2/modelvaralpha[i]))))^(1/2))
lowerfci[i]_1-(1/(1-g[i]))*(q[i]-
g[i]*(modelcov[i]/modelvaralpha[i])+(z[i]/abs(modelalpha[i]))
     *(modelvarbeta[i]-
(2*q[i]*modelcov[i])+((q[i]^2)*modelvaralpha[i])-
(g[i]*(modelvarbeta[i]-
     (modelcov[i]^2/modelvaralpha[i]))))^(1/2))
}

g9_g<1
lowerfcina_lowerfci[g9]
upperfcina_upperfci[g9]
widthfci_upperfcina-lowerfcina
not9_niter-length(widthfci)
coveragef_rep(0,length(widthfci))
for (j in 1:length(widthfci))
{
if (lowerfcina[j]<0.33 && upperfcina[j]>0.33) coveragef[j]_1
}

meanalpha_mean(modelalpha)
sealpha_sqrt(var(modelalpha))

meanbeta_mean(modelbeta)
sebeta_sqrt(var(modelbeta))

corrab_cor(modelalpha,modelbeta)
corrabcheck_mean(modelcov/(sqrt(modelvaralpha)*sqrt(modelvarbeta)))

meanp_mean(psim)
sep_sqrt(var(psim))

meannsigma_mean(nsigma)

meanwidthnci_mean(widthnci)
cicoveragen_sum(coveragen)/niter

meanwidthfci_mean(widthfci)
cicoveragef_sum(coveragef)/length(widthfci)

meanalpha
sealpha
meanbeta
sebeta
corrab
corrabcheck
meanp
sep
meannsigma
```

```
meanwidthnci
cicoveragen
not9
meanwidthfci
cicoveragef
```

B.2 CANCER SIMULATION CODE

```
niter_1000
alpha_c(rep(0,niter))
beta_c(rep(0,niter))
gamma_c(rep(0,niter))

alphaint_c(rep(0,niter))
betaint_c(rep(0,niter))
gammaint_c(rep(0,niter))
betagammaint_c(rep(0,niter))

survlambdauntrt_(-log(0.95)/5)
survlambdatrt_(-log(0.95)/5)

srlambdauntrt_3
srlambdatrt_3

recurrlambdauntrt_(-log(0.7)/3)
recurrlambdatrt_(-log(0.8)/3)

#simalpha_log(survlambdatrt/survlambdauntrt)
#simbeta_log(recurrlambdatrt/recurrlambdauntrt)
#simgamma_log(3)

for (j in 1:niter)
{
n0_500
n1_500
n_n0+n1

id_rep(1:n,1)
r_c(rep(0,n0),rep(1,n1))
wtk_rep(0,n)

start_rep(0,n)

timerecuntrt_c(rexp(n0,recurrlambdauntrt))
timerectrt_c(rexp(n1,recurrlambdatrt))
recurrtime_c(timerecuntrt,timerectrt)

timesurvuntrt_c(rexp(n0,survlambdauntrt))
timesurvtrt_c(rexp(n1,survlambdatrt))
survtime_c(timesurvuntrt,timesurvtrt)
```

```
survtime1_c(rep(0,length(survtime)))
timesr_c(rep(0,length(survtime)))

for (k in 1:length(survtime))
{  if (recurrtime[k]>survtime[k]) {survtime1[k]_survtime[k]}
      else if (recurrtime[k]<survtime[k] && r[k]==0)
            {timesr[k]_rexp(1,srlambdauntrt);
survtime1[k]_recurrtime[k]+timesr[k]}
      else if (recurrtime[k]<survtime[k] && r[k]==1)
            {timesr[k]_rexp(1,srlambdatrt);
survtime1[k]_recurrtime[k]+timesr[k]}
}

#censor_runif(n,0,quantile(survtime1,0.50))

status_c(rep(1,n))
#status[survtime<censor]_1

timedata_cbind(id,start,survtime1,recurrtime,status,r)

wtk[survtime1>recurrtime]_1
checkwt1_wtk==1
idtimedep_id[checkwt1]

id1_rep(0,n)
id2_rep(0,length(idtimedep))
r1_rep(0,n)
r2_rep(0,length(idtimedep))
start1_rep(0,length(start))
start2_rep(0,length(idtimedep))
stop1_rep(0,length(survtime1))
stop2_rep(0,length(idtimedep))
wt1_rep(0,length(wtk))
wt2_rep(0,length(idtimedep))
status1_rep(0,length(wtk))
status2_rep(0,length(idtimedep))

for (i in 1:length(id))
{
if (survtime[i]>recurrtime[i]) {id1[i]_id[i]; r1[i]_r[i]; start1[i]_start[i];
            stop1[i]_recurrtime[i]; wt1[i]_0; status1[i]_0;
            id2[i]_id[i]; r2[i]_r[i]; start2[i]_recurrtime[i];
            stop2[i]_survtime1[i]; wt2[i]_1; status2[i]_status[i]}
      else {id1[i]_id[i]; r1[i]_r[i]; start1[i]_start[i];
            stop1[i]_survtime1[i]; wt1[i]_0; status1[i]_status[i]}
}

idcheck_rep(0,n)
for (i in 1:n)
{
if (id2[i]!=0 && id2[i]!="NA") {idcheck[i]_1}
}


idfinal_c(id1,id2[idcheck==1])
rfinal_c(r1,r2[idcheck==1])
```

```
startfinal_c(start1,start2[idcheck==1])
stopfinal_c(stop1,stop2[idcheck==1])
wtfinal_c(wt1,wt2[idcheck==1])
statusfinal_c(status1,status2[idcheck==1])
final_cbind(idfinal,startfinal,stopfinal,wtfinal,statusfinal)

idcox_c(idfinal,idfinal)
r1cox_c(rfinal,rep(0,length(rfinal)))
r2cox_c(rep(0,length(rfinal)),rfinal)
etypecox_c(rep(1,length(idfinal)),rep(2,length(idfinal)))
wtcox_c(rep(0,length(idfinal)),wtfinal)

startcox_c(startfinal,startfinal)
stopcox_c(stopfinal,stopfinal)
statuscox_c(statusfinal,statusfinal)

model1_coxph(Surv(startcox,stopcox,statuscox)
      ~(r1cox + r2cox + wtcox)*strata(etypecox)+cluster(idcox))
alpha[j]_model1$coef[1]
beta[j]_model1$coef[2]
gamma[j]_model1$coef[3]
model2_coxph(Surv(startcox,stopcox,statuscox)
      ~(r1cox + r2cox + wtcox + r2cox*wtcox)*strata(etypecox)+cluster(idcox))
alphaint[j]_model2$coef[1]
betaint[j]_model2$coef[2]
gammaint[j]_model2$coef[3]
betagammaint[j]_model2$coef[4]
}

meanalpha_mean(alpha)
meanbeta_mean(beta)
meangamma_mean(gamma)

meanalphaint_mean(alphaint)
meanbetaint_mean(betaint)
meangammaint_mean(gammaint)
meanbetagammaint_mean(betagammaint)

#simalpha

meanalpha
meanbeta
meangamma

meanalphaint
meanbetaint
meangammaint
meanbetagammaint
```

B.3 QTWIST CODE

```
printci
decimals 2
file 'c03df.dat'
format free
variables 12
read
events <3,8><4,9><5,10><6,11><7,12>
names df1 df2 df3 df4 df5
timepoints 60
treatment 1
compare 1v0
summary
seed 745236472
sample 2000
means
utility 0.2 0.4 0.6 0.8 1
```

BIBLIOGRAPHY

1. Herson J; The use of surrogate endpoints in clinical trials (an introduction to a series of four papers). Stat Med 8: 403-404, 1989
2. Prentice RL; Surrogate endpoints in clinical trials: definition and operational criteria. Stat Med 8: 431-440, 1989
3. Fleming TR; Surrogate markers in AIDS and cancer trials. Stat Med 13: 1423-1435, 1994
4. De Gruttola V, Fleming TR, Lin DY, et al; Perspective: validating surrogate markers – are we being naïve?. J Infec Dis 175: 237-246, 1997
5. Fleming TR, Prentice RL, Pepe MS, et al; Surrogate and auxiliary endpoints in clinical trials, with potential applications in cancer and AIDS research. Stat Med 13: 955-968, 1994
6. Flandre P, O'Quigley J; A two-stage procedure for survival studies with surrogate endpoints. Biometrics 51: 969-976, 1995
7. Venkatraman ES, Begg CB; Properties of a nonparametric test for early comparison of treatments in clinical trials in the presence of surrogate endpoints. Biometrics 55: 1171-1176, 1999
8. Freedman LS, Graubard BL; Statistical validation of intermediate endpoints for chronic diseases. Stat Med 11: 167-178, 1992
9. Lin DY, Fleming TR, De Gruttola V; Estimating the proportion of treatment effect explained by a surrogate marker. Stat Med 16: 1515-1527, 1997
10. Wei LJ, Lin DY, Weissfeld L; Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. J Am Stat Assoc 84: 1065-1073, 1989
11. Lin DY; MULCOX: a computer program for the Cox regression analysis of multiple failure time variables. Comp Meth Prog Biomed 32: 125-135, 1990
12. Lin DY; MULCOX2: a general computer program for the Cox regression analysis of multivariate failure time data. Comp Meth Prog Biomed 40: 279-293, 1993
13. Sargent DJ, Wieand HS, Haller DG, et al; Disease free survival versus overall survival as a primary end point for adjuvant colon cancer studies: individual patient data from 20,898 patients on 18 randomized trials. J Clin Oncol 23: 8664-8670, 2005
14. Zamboni BA, Yothers G, Choi M, Fuller CD, Dignam JJ, Raich PC, Thomas CR, O'Connell MJ, Wolmark N, Wang SJ. Conditional Survival for Patients with Colon Cancer: An Analysis of NSABP Trials C-03 through C-07. Journal of Clinical Oncology. 28(15):2544-2548:2010.
15. Henson DE, Ries LA. On the estimation of survival. Semin Surg Oncol10:2-6, 1994
16. Gloeckler Ries LA, Reichman ME, Lewis DR, et al. Cancer survival and incidence from the Surveillance, Epidemiology, and End Results (SEER) program. Oncologist 8:541-552, 2003

17. Davis FG, McCarthy BJ, Freels S, et al. The conditional probability of survival of patients with primary malignant brain tumors: surveillance, epidemiology, and end results (SEER) data. Cancer 85:485-491,1999
18. Hwang SL, Yang YH, Lieu AS, et al. The conditional survival statistics for survivors with primary supratentorial astrocytic tumors. J Neurooncol 50:257-264, 2000
19. Lin CL, Lieu AS, Lee KS, et al. The conditional probabilities of survival in patients with anaplastic astrocytoma or glioblastoma multiforme. Surg Neurol 60:402-406; discussion, 406, 2003
20. Fuller CD, Wang SJ, Thomas CR, Jr., et al. Conditional survival in head and neck squamous cell carcinoma: Results from the SEER dataset 1973-1998. Cancer 109:1331-1343, 2007
21. Henson DE, Ries LA, Carriaga MT. Conditional survival of 56,268 patients with breast cancer. Cancer 76:237-242, 1995
22. Meng L, Maskarinec G, Lee J. Ethnicity and conditional breast cancer survival in Hawaii. J Clin Epidemiol 50:1289-1296, 1997
23. Christensen LH, Engholm G, Ceberg J, et al. Can the survival difference between breast cancer patients in Denmark and Sweden 1989 and 1994 be explained by patho-anatomical variables?--A population-based study. Eur J Cancer 40:1233-1243, 2004
24. Wang SJ, Luh JY, Fuller CD, et al. Impact of ethnicity on conditional survival of breast cancer patients: Analysis from the SEER database. 28th Annual San Antonio Breast Cancer Symposium, San Antonio, TX, December 8, 2005. Abstr # 3090
25. Wang SJ, Luh JY, Fuller CD, et al. Geographic variation in conditional survival for breast cancer patients: A SEER database analysis. 29th Annual San Antonio Breast Cancer Symposium, San Antonio, TX, December 8, 2006. Abstr # 6058
26. Merrill RM, Henson DE, Barnes M. Conditional survival among patients with carcinoma of the lung. Chest 116:697-703, 1999
27. Skuladottir H, Olsen JH. Conditional survival of patients with the four major histologic subgroups of lung cancer in Denmark. J Clin Oncol 21:3035-3040, 2003
28. Wang SJ, Fuller CD, Thomas CR, Jr. Ethnic disparities in conditional survival of patients with non-small cell lung cancer. J Thorac Oncol 2:180-190, 2007
29. Fuller CD, Wang SJ, Kachnic L, et al. Conditional survival of gallbladder adenocarcinoma treated with radiotherapy: Analysis from the SEER database. Int J Radiat Oncol Biol Phys 63:S274, 2005
30. Thomas CR, Fuller CD, Starling GD, Wang SJ. Conditional survival in gallbladder carcinoma: Results from the SEER 11 dataset. J Clin Oncol, 2006 ASCO Annual Meeting Proceedings 2006; 24:4130. Abstr #4130
31. Wang SJ, Emery R, Fuller CD,et al. Conditional survival in gastric cancer: A SEER database analysis. Gastric Cancer 10:153-158, 2007
32. Merrill RM, Henson DE, Ries LA. Conditional survival estimates in 34,963 patients with invasive carcinoma of the colon. Dis Colon Rectum 41:1097-1106, 1998
33. Wang S, Zamboni B, Wieand H, et al. Conditional survival for patients with colon cancer: An analysis of National Surgical Adjuvant Breast and Bowel Project (NSABP) trials C-03 through C-06. J Clin Oncol, 2006 ASCO Annual Meeting Proceedings 2006;24. Abstr # 6005
34. Wang SJ, Fuller CD, Emery R, et al. Conditional survival in rectal cancer: A SEER database analysis. Gastrointestinal Cancer Research 2007;1:84-89.

35. Luh J, Wang S, Fuller C, et al. A SEER database analysis of conditional survival for prostate cancer patients. J Clin Oncol, 2006 ASCO Annual Meeting Proceedings 2006;24: Abstr # 14506
36. Choi M, Fuller CD, Thomas CR, Jr., et al. Conditional survival in ovarian cancer: Results from the SEER dataset 1988-2001. Gynecol Oncol 109:203-209, 2008
37. Moller MB, Pedersen NT, Christensen BE. Conditional survival of patients with diffuse large B-cell lymphoma. Cancer 106:2165-2170, 2006
38. Klish DS, Wang SJ, Kimler BF, et al. Improved prognosis with time after diagnosis for patients with anal cancer. American Radium Society, 88th Annual Meeting, Wailea, HI, May 6, 2006
39. Kim E, Wang SJ, Fuller CD, er al. Conditional survival for patients with thymic carcinoma from the SEER database. American Thoracic Society International Conference, San Francisco, CA, May 20, 2007. Abstr #5420
40. Janssen-Heijnen ML, Houterman S, Lemmens VE, et al. Prognosis for long-term survivors of cancer. Ann Oncol 18:1408-1413, 2007
41. Kato I, Severson RK, Schwartz AG. Conditional median survival of patients with advanced carcinoma: surveillance, epidemiology, and end results data. Cancer 92:2211-2219, 2001
42. Wolmark N, Rockette H, Fisher B, et al. The benefit of leucovorin-modulated fluorouracil as postoperative adjuvant therapy for primary colon cancer: Results from National Surgical Adjuvant Breast and Bowel Project protocol C-03. J Clin Oncol 11:1879-1887, 1993
43. Wolmark N, Rockette H, Mamounas E, et al. Clinical trial to assess the relative efficacy of fluorouracil and leucovorin, fluorouracil and levamisole, and fluorouracil, leucovorin, and levamisole in patients with Dukes' B and C carcinoma of the colon: Results from National Surgical Adjuvant Breast and Bowel Project C-04. J Clin Oncol 17:3553-3559, 1999
44. Wolmark N, Bryant J, Smith R, et al. Adjuvant 5-fluorouracil and leucovorin with or without interferon alfa-2a in colon carcinoma: National Surgical Adjuvant Breast and Bowel Project Protocol C-05. J Natl Cancer Inst 90:1810-1816, 1998
45. Lembersky BC, Wieand HS, Petrelli NJ, et al. Oral uracil and tegafur plus leucovorin compared with intravenous fluorouracil and leucovorin in stage II and III carcinoma of the colon: Results from National Surgical Adjuvant Breast and Bowel Project Protocol C-06. J Clin Oncol 24:2059-2064, 2006
46. Kuebler JP, Wieand HS, O'Connell MJ, et al. Oxaliplatin combined with weekly bolus fluorouracil and leucovorin as surgical adjuvant chemotherapy for stage II and III colon cancer: Results from NSABP C-07. J Clin Oncol 25:2198-2204, 2007
47. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER 17 Regs Limited-Use, Nov 2006 Sub (1973-2004 varying), National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2007, based on the November 2006 submission. http://canques.seer.cancer.gov/surv2004/Canques.cfg Accessed: 7-30-08
48. Bryant J, Fisher B, Gündüz N, Costantino JP, Emir B. S-phase fraction combined with other patient and tumor characteristics for the prognosis of node-negative, estrogen-receptor-positive breast cancer. Breast Cancer Res Treat 51:239-53, 1998
49. Gill S, Sargent D. End points for adjuvant therapy trials: Has the time come to accept disease-free survival as a surrogate end point for overall survival? Oncologist11:624-629, 2006

50. Harshman, Lauren C., et al. "Conditional survival of patients with metastatic renal-cell carcinoma treated with VEGF-targeted therapy: a population-based study." *The lancet oncology* 13.9 (2012): 927-935.
51. Xue, Q. Yu, Peter D. Baade, and Dianne L. O'Connell. "Conditional survival of cancer patients: an Australian perspective." *BMC cancer* 12.1 (2012): 460.
52. Schoop, Rotraut, Martin Schumacher, and Erika Graf. "Measures of prediction error for survival data with longitudinal covariates." *Biometrical Journal* 53.2 (2011): 275-293.
53. Wang, Samuel J., et al. "Individualized estimation of conditional survival for patients with head and neck cancer." *Otolaryngology--Head and Neck Surgery* 145.1 (2011): 71-73.
54. Wang, Samuel J., et al. "An interactive tool for individualized estimation of conditional survival in rectal cancer." *Annals of surgical oncology* 18.6 (2011): 1547-1552.
55. Dasgupta, Paramita, Danny R. Youlden, and Peter D. Baade. "Multiple primary cancers among colorectal cancer survivors in Queensland, Australia, 1996–2007." *Cancer Causes & Control* 23.8 (2012): 1387-1398.
56. Tan, Marcus CB, et al. "Prognostic significance of early recurrence: a conditional survival analysis in patients with resected colorectal liver metastasis." *HPB* 15.10 (2013): 803-813.
57. Nicolaie, M. A., et al. "Dynamic prediction by landmarking in competing risks." *Statistics in medicine* 32.12 (2013): 2031-2047.
58. Dasgupta, Paramita, Danny R. Youlden, and Peter D. Baade. "An analysis of competing mortality risks among colorectal cancer survivors in Queensland, 1996–2009." *Cancer Causes & Control* 24.5 (2013): 897-909.
59. Shikanov, Sergey, and Scott E. Eggener. "Hazard of prostate cancer specific mortality after radical prostatectomy." *The Journal of urology* 187.1 (2012): 124-128.
60. Nicolaie, M. A., et al. "Dynamic Pseudo-Observations: A Robust Approach to Dynamic Prediction in Competing Risks." *Biometrics* 69.4 (2013): 1043-1052.
61. Turner, Donna. "Surviving cancer." (2011).
62. Zhu, Y., et al. "Conditional survival in patients treated with vascular endothelial growth factor-targeted therapy for advanced renal cell carcinoma." *Journal of cancer research and clinical oncology* 138.11 (2012): 1901-1907.
63. Mishra, Mark V., et al. "Conditional Survival Probabilities for Patients With Resected Pancreatic Adenocarcinoma." *IIB* 2602 (2012): 53-29.
64. Gelber, Richard D., et al. "Comparing treatments using quality-adjusted survival: The Q-TWiST method." *The American Statistician* 49.2 (1995): 161-169.
65. Gordon, Alan N., et al. "Recurrent epithelial ovarian carcinoma: a randomized phase III study of pegylated liposomal doxorubicin versus topotecan." *Journal of Clinical Oncology* 19.14 (2001): 3312-3322.
66. Blazeby, Jane M., et al. "A prospective longitudinal study examining the quality of life of patients with esophageal carcinoma." *Cancer* 88.8 (2000): 1781-1787.
67. Cole, Bernard F., et al. "Quality-of-life-adjusted survival analysis of interferon alfa-2b adjuvant treatment of high-risk resected cutaneous melanoma: an Eastern Cooperative Oncology Group study." *Journal of Clinical Oncology* 14.10 (1996): 2666-2673.
68. Keller, Alan M., et al. "Randomized phase III trial of pegylated liposomal doxorubicin versus vinorelbine or mitomycin C plus vinblastine in women with taxane-refractory advanced breast cancer." *Journal of Clinical Oncology* 22.19 (2004): 3893-3901.

69. Gelber, R. D., et al. "Adjuvant chemotherapy plus tamoxifen compared with tamoxifen alone for postmenopausal breast cancer: meta-analysis of quality-adjusted survival." *The Lancet* 347.9008 (1996): 1066-1071.
70. Gelber, Richard D., et al. "A quality-adjusted time without symptoms or toxicity (Q-TWiST) analysis of adjuvant radiation therapy and chemotherapy for resectable rectal cancer." *Journal of the National Cancer Institute* 88.15 (1996): 1039-1045.
71. Schwartz, Carolyn E., et al. "The quality-of-life effects of interferon beta-1b in multiple sclerosis: an extended Q-TWIST analysis." *Archives of neurology* 54.12 (1997): 1475-1480.
72. Cole, Bernard F., et al. "Quality-of-life–adjusted survival comparison of sustained-release cytosine arabinoside versus intrathecal methotrexate for treatment of solid tumor neoplastic meningitis." *Cancer* 97.12 (2003): 3053-3060.
73. Fairclough, Diane L., et al. "Quality of life and quality adjusted survival for breast cancer patients receiving adjuvant therapy." *Quality of Life Research* 8.8 (1999): 723-731.
74. Zee, B., et al. "Quality-adjusted time without symptoms or toxicity analysis of interferon maintenance in multiple myeloma." *Journal of Clinical Oncology* 16.8 (1998): 2834-2839.
75. Parsons, S. K., et al. "Quality-adjusted survival after treatment for acute myeloid leukemia in childhood: a Q-TWiST analysis of the Pediatric Oncology Group Study 8821." *Journal of Clinical Oncology* 17.7 (1999): 2144-2144.
76. Schwartz, Carolyn E., Bernard F. Cole, and Richard D. Gelber. "Measuring patient-centered outcomes in neurologic disease: extending the Q-TWiST method." *Archives of neurology* 52.8 (1995): 754-762.
77. Levy, Vincent, et al. "Evaluating treatment strategies in chronic lymphocytic leukemia: use of quality-adjusted survival analysis." *Journal of clinical epidemiology* 54.7 (2001): 747-754.
78. Cole, Bernard F., et al. "Quality-of-life-adjusted survival analysis of interferon alfa-2b treatment for advanced follicular lymphoma: an aid to clinical decision making." *Journal of clinical oncology* 16.7 (1998): 2339-2344.
79. Adler, Frederick R., et al. "Lung transplantation for cystic fibrosis." *Proceedings of the American Thoracic Society* 6.8 (2009): 619-633.
80. Zhao, Yufan, et al. "Reinforcement learning strategies for clinical trials in nonsmall cell lung cancer." *Biometrics* 67.4 (2011): 1422-1433.
81. Borie, Frédéric, et al. "Cost-effectiveness of two follow-up strategies for curative resection of colorectal cancer: comparative study using a Markov model." *World journal of surgery* 28.6 (2004): 563-569.
82. Schwartz, C. E., et al. "The Q-TWiST approach to assessing health-related quality of life in epilepsy." *Quality of Life Research* 4.2 (1995): 135-141.
83. Hurny, C., et al. "Subjective health estimations (SHE) in patients with advanced breast cancer: an adapted utility concept for clinical trials." *British journal of cancer* 77.6 (1998): 985.
84. Cella, David, et al. "Role of quality of life in men with metastatic hormone-refractory prostate cancer: how does atrasentan influence quality of life?." *European urology* 49.5 (2006): 781-789.
85. Hornberger, John, et al. "The cost-effectiveness of bortezomib in relapsed/refractory multiple myeloma: Swedish perspective." *European journal of haematology* 85.6 (2010): 484-491.

86. Porcher, R., et al. "Evaluating high dose therapy in multiple myeloma: use of quality-adjusted survival analysis." *Quality of Life Research* 11.2 (2002): 91-99.
87. Zbrozek, Arthur S., et al. "Q-TWiST analysis of patients receiving temsirolimus or interferon alpha for treatment of advanced renal cell carcinoma." *Pharmacoeconomics* 28.7 (2010): 577-584.
88. Corey-Lisle, Patricia K., et al. "Q-TWiST analysis of ixabepilone in combination with capecitabine on quality of life in patients with metastatic breast cancer." *Cancer* 118.2 (2012): 461-468.
89. Lehmann, Jan, Margitta Retz, and Michael Stöckle. "Is there standard chemotherapy for metastatic bladder cancer? Quality of life and medical resources utilization based on largest to date randomized trial." *Critical reviews in oncology/hematology* 47.2 (2003): 171-179.
90. Wang, J., et al. "A Q-TWiST analysis comparing panitumumab plus best supportive care (BSC) with BSC alone in patients with wild-type KRAS metastatic colorectal cancer." *British journal of cancer* 104.12 (2011): 1848-1853.
91. Konski, Andre A., et al. "Quality-adjusted survival analysis of Radiation Therapy Oncology Group (RTOG) 90-03: Phase III randomized study comparing altered fractionation to standard fractionation radiotherapy for locally advanced head and neck squamous cell carcinoma." *Head & neck* 31.2 (2009): 207-212.
92. Radice, Davide, and Alberto Redaelli. "Q-TWiST analysis of cyclophosphamide, epirubicin, fluorouracil versus cyclophosphamide, methotrexate, fluorouracil treatment for premenopausal women with node-positive breast cancer." *Pharmacoeconomics* 23.1 (2005): 69-75.
93. Miller, Dave P., et al. "Quality-Days Incrementally Gained: A New Approach for Short-Term Evaluations Based on the Extended Q-Twist Method Illustrated in a Clinical Study of Reflux Disease Patients*." *Drug information journal* 34.2 (2000): 469-481.
94. Tiong, Albert, and June Corry. "The role of palliative care in head and neck cancer." *Radiation Oncology in Palliative Cancer Care* (2013): 126-144.
95. Cella, David, and Carol M. Moinpour. "Quality of Life and Quality Adjusted Survival for Breast Cancer Patients Receiving Adjuvant Therapy." *Quality of Life Research* (1999): 723-731.
96. Zhao, Hongwei, and Anastasios A. Tsiatis. "Efficient Estimation of the Distribution of Quality-Adjusted Survival Time." *Biometrics* 55.4 (1999): 1101-1107.
97. Zhao, Hongwei, and Anastasios A. Tsiatis. "Estimating mean quality adjusted lifetime with censored data." *Sankhyā: The Indian Journal of Statistics, Series B* (2000): 175-188.
98. Wang, Hongkun, and Hongwei Zhao. "Regression analysis of mean quality-adjusted lifetime with censored data." *Biostatistics* 8.2 (2007): 368-382.
99. Wang, Hongkun, and Yichuan Zhao. "A comparison of some confidence intervals for the mean quality-adjusted lifetime with censored data." *Computational Statistics & Data Analysis* 53.7 (2009): 2733-2739.
100. Raboud, J. M., et al. "Estimating the effect of treatment on quality of life in the presence of missing data due to drop-out and death." *Quality of Life Research* 7.6 (1998): 487-494.
101. Andrei, Adin-Cristian, and Susan Murray. "Regression Models for the Mean of the Quality-of-Life-Adjusted Restricted Survival Time Using Pseudo-Observations." *Biometrics* 63.2 (2007): 398-404.

102. Fine, Jason P., and Richard D. Gelber. "Joint Regression Analysis of Survival and Quality-Adjusted Survival." *Biometrics* 57.2 (2001): 376-382.
103. Bonetti, Marco, Bernard F. Cole, and Richard D. Gelber. "A method-of-moments estimation procedure for categorical quality-of-life data with nonignorable missingness." *Journal of the American Statistical Association* 94.448 (1999): 1025-1034.
104. Almanassra, Mahmoud, Jiantian Wang, and S. Jeyaratnam. "Monotonic estimators for the survival function of quality adjusted lifetime." *Communications in Statistics—Theory and Methods* 34.5 (2005): 1217-1231.
105. Revicki, Dennis A., et al. "Analyzing Oncology Clinical Trial Data Using the Q-TWiST Method: Clinical Importance and Sources for Health State Preference Data*." *Quality of life research* 15.3 (2006): 411-423.
106. Whately-Smith, Caroline, et al. "Utility values in health technology assessments: a statistician's perspective." *Pharmaceutical statistics* 13.3 (2014): 184-195.
107. Tate, Wendy R., and Grant H. Skrepnek. "Quality-adjusted time without symptoms or toxicity (Q-TWiST): patient-reported outcome or mathematical model? A systematic review in cancer." *Psycho-Oncology* (2014).
108. Cole, Bernard F. "QAS 2.0 Usage Guide."