# CENTER FOR DRUG EVALUATION AND RESEARCH

## APPLICATION NUMBER: NDA 20-571/S-008

## STATISTICAL REVIEW(S)

150 GUINN

**Statistical Review and Evaluation**

AUG 3 1998

| | |
|---|---|
| **NDA Number:** | 20,571 suppl. SE1 |
| **Applicant:** | Pharmacia & Upjohn Co. |
| **Name of Drug:** | CPT-11 (Camptosar, Irinotecan Hydrochloride Injection) |
| **Indication:** | Treatment of metastatic colorectal cancer after failing 5FU therapy. |
| **Documents Reviewed:** | Vols. 1.73-1.75, 1.80-1.82, 1.102-1.106, 4.001, S17.1 dated 17 Apr 1998 |
| **Medical Reviewer:** | Isagani Chico, M.D. |
| **Statistical Reviewer:** | David Smith, Ph.D. |

## 1. Background and Overview

In order to support labeling for the indication of treatment of metastatic colorectal cancer after failing a 5FU regimen, the sponsor submitted a supplemental NDA which is comprised of two Phase III trials and four Phase II trials. The sponsor's submission included the reports of the two pivotal Phase III studies and the four supportive Phase II studies. We will only consider the pivotal studies in this review.

A brief summary of the pivotal studies appears below.

| Study | Type | N | Arms |
|---|---|---|---|
| | Randomized Ph III | 176 | CPT-11 + BSC* |
| | | 88 | BSC |
| | Randomized Ph III | 127 | CPT-11 |
| | | 129 | Best chemotherapy regimen |

* "BSC" herein stands for "best supportive care."

The next section includes relevant statistical issues for these studies. The following sections will discuss these studies in more detail and will follow the following format:

1. General description of study
2. Efficacy endpoints and results
3. Summary and conclusions

The last two sections will include an integrated summary of efficacy and overall conclusions and recommendations for the submission.

References will follow the review.

## 2. Statistical Issues

- The following time to event endpoints were not specified in either the protocols: symptom-free duration in patients who were asymptomatic at baseline, time until pain in patients pain-free at baseline, time until performance status deterioration, time until weight loss. These endpoints should be considered as secondary since the analysis was retrospective.
- As a secondary analysis for studies Cox modeling for survival was performed and variables were selected using a stepwise selection procedure. Some of the variables that were considered for the Cox regressions were originally numerical variables that were classified into subgroups, creating categorical (indicator) variables. However, the classifications seemed to be based on subdivisions that were not necessarily clinically relevant, but ones that distributed the patients equally throughout the number of classes. Therefore, variables that were determined to be statistically significant in the Cox regressions may not have any clinical interpretation because the indicator variables are not based on clinically relevant intervals.
- The quality of life analysis consisted of comparing subscale measurements at each follow up to baseline measurements using MANOVA or ANOVA techniques. However, conclusions based on these methods may be biased, due to the presence of missing data. The sponsor did not perform a longitudinal quality of life analysis, which is more appropriate in determining trends over time and variance inflation due to patient dropout. There was also no attempt to adjust Type I error for the multiplicity of subscales considered.

## 3. Pivotal Phase III Trials

### 3.1 Description of Study

**Study Objective:** To evaluate survival of patients with metastatic colorectal cancer that have failed previous 5FU therapy.

**Study Enrollment Period:** September 1995 - March 1996

**Study Design:** Open label, multi-center randomized Phase III study. The only stratification factor was center. There may have been a slight imbalance in performance status at baseline. The sponsor's test for imbalance of performance status was statistically significant at 0.02. The p-value for Fisher's exact test was 0.060 for performance status of 0 or 1 versus 2 between the arms.

**Sample Size:** Out of 279 patients randomized, 189 were assigned to the CPT-11 + BSC arm and 90 were assigned to the BSC-only arm.

The sponsor assumed that the one-year survival was 35% for the CPT-11 + BSC group and 20% for the BSC-only group. Under a two-sided logrank with a Type I error of 0.05 and power of 0.80, 264 patients (176 in the CPT-11 + BSC arm and 88 in the BSC-only arm) would detect a significant difference between these one-year survival rates, assuming a 2:1 randomization among the two arms, a 5% loss-to-follow up, and a minimum follow up of 6 months.

**Interim Analysis:** No interim analysis was specified.

**Dose:** The CPT-11 + BSC arm received 350 mg/m$^2$ as a 90-minute intravenous on day 1 and this was repeated every 21 days. Best supportive care varied among centers. See the FDA medical review of CPT-11 for further details.

**Criteria for Evaluation:** The primary efficacy endpoint was survival on the intent-to-treat population. The prognostic factors specified in the protocol were resistance to prior 5FU therapy, duration of prior 5FU therapy, age, performance status, visceral involvement, number of metastatic sites by organ, intent of and response to prior chemotherapy.

Aditional prognostic factors were added into the survival analysis to take into account recent findings in the literature. These appear in Figure 1.

**Figure 1.** Prognostic variables for a Cox regression analysis on survival for study

- sex
- weight loss three months before baseline
- hepatic metastases at inclusion
- site of colorectal cancer (colon right vs. other)
- white blood cells counts
- hemoglobin
- platelets
- number of organs involved
- percent of the following biochemical parameters in the upper normal range at baseline: LDH, SGOT, SGPT, bilirubin, alkaline phosphatase, total protein
- CEA value at baseline

All prognostic factors were studied retrospectively using stratified logrank analysis and Cox models. A stepwise Cox regression procedure was performed only on those variables that were significant at 0.10 in a univariate model.

The sponsor tested for baseline imbalances due to randomization for several patient characteristics. All but one characteristic were non-significant. The sponsor's test for imbalance of performance status was significant at 0.02.

**Efficacy Endpoints**

Table 3.1 shows a summary of the time to event efficacy endpoints in study

**Table 3.1.** Analysis of time to event endpoints for the intent-to-treat population of Study Numbers in parentheses are the 95% confidence interval for the point estimate.

| Endpoint | Arm | Median (months) | p-value |
|---|---|---|---|
| Survival (Primary Endpoint) | CPT-11 + BSC | 9.2 (8.4-10.7) | 0.0001 |
| | BSC-only | 6.5 (5.0-7.6) | |
| Survival calculated from date of diagnosis | CPT-11 + BSC | 33.8 (28.7-36.6) | 0.0076 |
| | BSC-only | 26.5 (21.6-31.6) | |
| Symptom-Free Surv in Pats. Asymp. at Baseline | CPT-11 + BSC | 5.9 (3.8-7.6) | 0.2049 |
| | BSC-only | 4.1 (2.2-6.9) | |
| Time Until Pain in Pain-Free Pats. at Baseline | CPT-11 +BSC | 6.9 (5.8-8.4) | 0.0026 |
| | BSC-only | 2.0 (1.8-5.1) | |
| Time to Performance Status Deterioration | CPT-11 + BSC | 5.7 (4.3-6.6) | 0.0001 |
| | BSC-only | 3.3 (1.9-3.7) | |
| Time to Wt. Loss > 5% of Wt. At Baseline | CPT-11 + BSC | 6.4 (5.5-7.7) | 0.0183 |
| | BSC-only | 4.2 (3.4-5.1) | |

**Survival**

Survival was defined as the time from the date of randomization until death due to any cause or loss to follow-up. Patients who had not died were censored at their last follow-up or the cutoff date. For the CPT-11 + BSC arm, the median survival was 9.2 months (95% CI: 8.4-10.7 months) and for the BSC-only arm, the median survival was 6.5 months (95% CI: 5.0-7.6 months). The unadjusted logrank p-value for survival was 0.0001, which is statistically significant.

As an exploratory analysis, this reviewer examined survival if it were to be calculated as the time from cancer diagnosis until death or censoring. For the CPT-11 + BSC arm and the BSC-only arm, the median survival under this definition was 33.8 months (95% CI: 28.7-36.6 months) and 26.5 months (95% CI: 21.6-31.6 months) respectively. The unadjusted logrank p-value for this difference in survival was 0.0076, which is statistically significant.

**Symptom-Free Survival**

Symptom free survival in patients asymptomatic at baseline was defined as the time from the date of randomization until date of first appearance of a symptom. Tumor related symptoms were examined, as opposed to drug-related symptoms; however, if all symptoms were taken into account, too few patients would have remained symptom-free at baseline. Therefore, the sponsor chose the most commonly-occurring symptoms during the study. The symptoms that were taken into account were anorexia, asthenia, constipation, diarrhea, fever in absence of neutropenia, hemorrhage, symptoms of liver disorders, nausea, vomiting, pain and abdominal pain. Patients were censored if the patient had

not died and no symptoms appeared before the cutoff date. There were 59 patients in the CPT-11 + BSC arm and 19 patients in the BSC-only arm who were asymptomatic at baseline. For the CPT-11 + BSC arm and the BSC-only arm, the median symptom free survival was 5.9 months (95% CI: 3.8-7.6 months) and 4.1 months (95% CI: 2.2-6.9 months) respectively. The wide confidence intervals are partly due to the small number of patients included in the analysis. The unadjusted logrank p-value was 0.2049.

**Time Until Worsening of Pain**

Pain free survival in patients pain free at baseline was defined as the time from the date of randomization until date of first appearance of a symptom. Patients were censored if the patient had not died and no symptoms appeared before the cutoff date. For the CPT-11 + BSC arm and the BSC-only arm, the median time to performance status deterioration was 6.9 months (95% CI: 5.8-8.4 months) and 2.0 months (95% CI: 1.8-5.1 months) respectively. The unadjusted logrank p-value was 0.0026, which is statistically significant.

Among CPT-11 + BSC patients who did not take analgesics at baseline, 13.5% were taking opioids and 17.6% were taking non-opioids at week 20. Among BSC-only patients who did not take analgesics at baseline, 33.3% were taking opioids and 28.6% were taking non-opioids at week 20. For those patients who were not taking opioids at baseline, 16.8% of patients on the CPT-11 + BSC arm were taking opioids at 20 weeks, compared to 38.5% of patients on the BSC-only arm.

**Time to Performance Status Deterioration**

Time to definitive WHO performance status deterioration was defined as the time from the date of randomization until definitive WHO performance status deterioration. Patients were censored if the patient had not died or experienced WHO performance status deterioration. Patients were also censored if their performance status evaluations ceased. Out of 99 patients in the CPT-11 + BSC arm who had a performance status greater than 0 at baseline, thirty-three patients, or 33.3%, had improved performance status. Out of the 62 patients in the BSC-only arm who had a performance status greater than 0 at baseline, 7 patients, or 11.3%, had improved performance status. For the CPT-11 + BSC arm and the BSC-only arm, the median time to performance status deterioration was 5.7 months (95% CI: 4.3-6.6 months) and 3.3 months (95% CI: 1.9-3.7 months) respectively. The unadjusted logrank p-value was 0.0001, which is statistically significant.

**Time to Weight Loss Greater Than Five Percent**

Time to > 5% weight loss was defined as the time from the date of randomization until date in which the patient lost > 5% of measured weight at baseline. Patients were censored if the patient had not died or experienced > 5% weight loss at the cutoff date. Patients were also censored if their weight measurements ceased. For the CPT-11 + BSC arm and the BSC-only arm, the median time to performance status deterioration was 6.4 months (95% CI: 5.5-7.7 months) and 4.2 months (95% CI: 3.4-5.1 months) respectively. The unadjusted logrank p-value was 0.0183, which is statistically significant.

**Cox Analysis of Survival**

As a secondary analysis, a Cox model adjusted for a group of prognostic factors was fitted for survival. Some of the prognostic factors were specified in the protocol, while others were additionally added in order to take into account recent developments in the medical literature. The clinical prognostic factors appear in Figure 1. A Cox univariate analysis was performed to select the significant variables for a multiple Cox regression. Variables were selected for the multiple regression if they were statistically significant in a univariate model using an alpha of 0.10. The variables that were significant at 0.10 in their respective univariate models were performance status, organ

involvement, liver mets, tumor location (colon right vs. other), weight loss, platelets, hemoglobin, white blood cells, alkaline phosphatase, and bilirubin.

In the multiple regression on these 10 variables, a stepwise method was chosen, where the p-value to enter the model was 0.10 and the p-value to stay in the model was 0.05. The results of this model appear in Table 3.2.

**Table 3.2.** Results of stepwise Cox regression on survival using the prognostic factors in Figure 1. Treatment group was purposely withheld from this model.

| Prognostic Factor | Risk Ratio | p-value |
|---|---|---|
| WHO performance status (0,1 vs. >1) | 2.168 | <0.001 |
| Number of organs involved | 1.471 | 0.017 |
| Liver metastases | 1.592 | 0.032 |
| Weight loss (≥ 5% vs. < 5%) | 3.550 | <0.001 |
| Hemoglobin (g/dl) | 2.162 | <0.001 |
| Alk. Phosphatase (% of ULN) | 1.562 | 0.016 |

When one adds the additional variable of treatment group into this model, the treatment group continues to be significant (risk ratio = 1.711; p-value = 0.001).

The sponsor classified numeric variables such as hemoglobin, alkaline phosphatase, and others into indicator variables. It was not clear to this reviewer as to how the cutoff points for each subdivision were determined, although it appears as if they were chosen for nearly-equal distribution of patients across the selected number of intervals. See Table 3.3 for examples of the subdivisions of the indicator variables.

**Table 3.3.** Subdivisions of variables considered in the Cox models in V301. The units for each variable were not specified in the subdivisions.

| Platelets | | WBC | | SGPT | | SGOT | | Alk. Phos. | |
|---|---|---|---|---|---|---|---|---|---|
| Subdiv. | N | Subdiv. | N | Subdiv. | N | Subdiv. | N | Subdiv. | N |
| [0, 222) | 67 | [0.0, 6.3) | 66 | [0, 35) | 63 | [0, 56) | 64 | [0, 73) | 68 |
| [222, 277) | 72 | [6.3, 7.6) | 67 | [35, 54) | 65 | [56, 76) | 62 | [73, 107) | 68 |
| [277, 341) | 70 | [7.6, 9.5) | 75 | [54, 89) | 64 | [76, 116) | 64 | [107, 167) | 69 |
| [341, ∞) | 70 | [9.5, ∞) | 71 | [89, ∞) | 64 | [116, ∞) | 64 | [167, ∞) | 69 |

Although treatment arm was significant in the final Cox model for survival, it is this reviewer's opinion that the significance for hemoglobin and alkaline phosphatase in Table 3.2 should be interpreted with caution since their categorical analog may not be based on clinically relevant cutoff points.

This reviewer consulted with the medical reviewer and performed an exploratory analysis for those variables with clear, clinically relevant subdivisions. Subjectively dividing numerical variables without any type of standard may bias the regression estimates and lead to spurious conclusions about a variable's influence on the endpoint. For this reason, the covariates that were considered in this model were the ones that had clear clinical demarcations. Table 3.4 shows the covariates considered as well as the cutoffs used for a multiple Cox regression. The variable selection method for this regression analysis was the same as the sponsor's; namely, the variables that were significant at 0.10 in their respective univariate models were eligible for the final stepwise selection procedure. We also performed a backward elimination selection procedure (p-value of 0.05 required for retention) and a forward selection procedure (p-value of 0.10 required for entry) to investigate the consistency of our model-building.

**Table 3.4.** Variables and cutoff values for an exploratory Cox regression for survival in Study The "P-val." column refers to the p-value for the univariate Cox regression.

| Variable | Cutoffs | P-val. |
|---|---|---|
| Age | 0 to 64 vs. 65 and older | 0.942 |
| Gender | Male vs. female | 0.372 |
| Performance Status | 0 or 1 vs. 2 | 0.001 |
| Num. Of Organs Involved | 1 to 3 vs. more than 3 | 0.166 |
| Intent of prior chemo. | Adjuv. vs. advanced or adv. adjuv. | 0.567 |
| Best Resp. on 5FU | CR or PR vs. SD or PD | 0.515 |
| Liver metastases | Yes vs. no | 0.004 |
| Tumor location | colon right vs. other | 0.006 |
| Type of last prior chemo. | 5FU bolus vs. non-bolus | 0.709 |

The univariate regression p-values appear in the rightmost column Table 3.4. The variables that were eligible for the selection procedure were performance status, liver metastases, and tumor location, which is a different subset of variables from those selected by the sponsor.

The results of the variable selection procedures appear in Table 3.5.

**Table 3.5.** The results of three variable selection techniques using the variables in Table 3.4 in a Cox regression on survival. "Vars. selected" indicated the variables selected in the final model and "+Treatment" indicates the p-value when treatment arm is included in the final model.

| Stepwise | | Forward | | Backward | |
|---|---|---|---|---|---|
| Vars. selected | p-val | Vars. selected | p-val | Vars. selected | p-val |
| Liver mets.<br>Perf stat. | 0.003<br>0.001 | Liver mets.<br>Perf stat. | 0.003<br>0.001 | Liver mets.<br>Perf. stat. | 0.003<br>0.001 |
| +Treatment | 0.001 | +Treatment | 0.001 | +Treatment | 0.001 |

Note that the stepwise procedure selects the same model as the forward or backward procedures. Also, treatment arm is statistically significant in each of the final models.

**Quality of Life**

Quality of life (QOL) for study will be considered in Section 4.

### 3.2 Description of Study

**Study Objective:** To evaluate survival of patients with metastatic colorectal cancer that have failed previous 5FU therapy.

**Study Enrollment Period:** September 1995 - June 1996

**Study Design:** Open label, multi-center randomized Phase III study. The stratification factors were center and performance status.

**Sample Size:** Out of 258 patients randomized, 129 were assigned to the CPT-11 single-agent arm and 129 were assigned to the "best chemotherapy available" group.

The sponsor assumed that the one-year survival was 35% for the CPT-11 group and 20% for the chemo. group. Under a two-sided logrank with a Type I error of 0.05 and power of 0.80, 264 patients (129 in the CPT-11 arm and 129 in the BSC-only arm) would detect a significant difference between these one-year survival rates, assuming a 1:1 randomization among the two arms, a 5% loss-to-follow up, and a minimum follow up of 9 months and 7.5 months for the CPT-11 and chemotherapy group, respectively. Since this calculation of the number of patients assumes that accrual follows a uniform probability distribution but accrual was slower at the beginning of the recruitment than at the end, the follow-up period was extended until the 177$^{th}$ death to have the required power of 80%.

**Interim Analysis:** No interim analysis was specified.

**Dose:** The CPT-11 arm received 350 mg/m$^2$ as a 90-minute intravenous on day 1 and this was repeated every 21 days. Best chemotherapy varied among centers. See the FDA medical review of CPT-11 for further details.

**Criteria for Evaluation:** The primary efficacy endpoint was survival on the intent-to-treat population. The prognostic factors specified in the protocol were resistance to prior 5FU therapy, duration of prior 5FU therapy, age, performance status, visceral involvement, number of metastatic sites by organ, intent of and response to prior chemotherapy.

All prognostic factors were studied retrospectively using stratified logrank analysis and Cox models. A stepwise Cox regression procedure was performed only on those variables that were significant at 0.10 in a univariate model. The set of all prognostic factors appears in Figure 2.

**Figure 2.** Prognostic variables for a Cox regression analysis on survival for study

- sex
- weight loss three months before baseline
- hepatic metastases at inclusion
- site of colorectal cancer (colon right vs. other)
- white blood cells counts
- hemoglobin
- platelets
- number of organs involved
- percent of the following biochemical parameters in the upper normal range at baseline: LDH, SGOT, SGPT, bilirubin, alkaline phosphatase, total protein
- CEA value at baseline

**Efficacy Endpoints**

Table 3.6 shows a summary of the time to event efficacy endpoints in study

**Table 3.6.** Analysis of time to event endpoints for the intent-to-treat population of Study
Numbers in parentheses are the 95% confidence interval for the point estimate.

| Endpoint | Arm | Median (months) | p-value |
|---|---|---|---|
| Survival (Primary Endpoint) | CPT-11 | 10.8 (9.5-12.8) | 0.0351 |
| | Chemotherapy | 8.5 (7.7-10.5) | |
| Survival calculated from date of diagnosis | CPT-11 | 33.6 (28.3-40.4) | 0.0073 |
| | Chemotherapy | 27.0 (23.9-30.7) | |
| Progression Free Survival | CPT-11 | 4.2 (3.8-4.8) | 0.0295 |
| | Chemotherapy | 2.9 (2.6-3.7) | |
| Symptom-Free Surv. in Pats. Asymp. at Baseline | CPT-11 | 8.1 (6.1-10.7) | 0.2303 |
| | Chemotherapy | 7.0 (4.4-8.7) | |
| Time Until Pain in Pain-Free Pats. at Baseline | CPT-11 | 10.3 (7.8-BE*) | 0.0586 |
| | Chemotherapy | 8.5 (6.2-10.2) | |
| Time to Performance Status Deterioration | CPT-11 | 6.4 (5.2-7.6) | 0.1865 |
| | Chemotherapy | 5.1 (4.2-6.2) | |
| Time to Wt. Loss > 5% of Wt. At Baseline | CPT-11 | 8.9 (6.7-12.3) | 0.2279 |
| | Chemotherapy | 7.4 (4.7-11.6) | |

*BE stands for "biased estimate." The bias is due to a large number of censored observations.

**Survival**

For the CPT-11 + BSC arm, the median survival was 10.8 months (95% CI: 9.5-12.8 months) and for the BSC-only arm, the median survival was 8.5 months (95% CI: 7.7-10.5 months). The unadjusted logrank p-value for survival is 0.0351, which is statistically significant.

As an exploratory analysis, this reviewer examined survival if it were to be calculated as the time from cancer diagnosis until death or censoring. For the CPT-11 + BSC arm and the BSC-only arm, the median survival under this definition was 33.6 months (95% CI: 28.3-40.4 months) and 27.0 months (95% CI: 23.9-30.7 months) respectively. The unadjusted logrank p-value for this difference in survival is 0.0073, which is statistically significant.

**Progression free survival**

For the CPT-11 + BSC arm and the BSC-only arm, the median progression free survival was 4.2 months (95% CI: 3.8-4.8 months) and 2.9 months (95% CI: 2.6-3.7 months) respectively. The unadjusted logrank p-value for this difference in progression free survival is 0.0295, which is statistically significant.

**Symptom-Free Survival**

Symptom free survival in patients asymptomatic at baseline was defined as the time from the date of randomization until date of first appearance of a symptom. Tumor related symptoms were examined, as opposed to drug-related symptoms; however, if all symptoms were taken into account, too few patients would have remained symptom-free at baseline. Therefore, the sponsor chose the most commonly-occurring symptoms during the study. The symptoms that were taken into account were

anorexia, asthenia, constipation, diarrhea, fever in absence of infection and neutropenia, hemorrhage, symptoms of liver disorders, nausea, vomiting, pain and abdominal pain. Patients were censored if the patient had not died and no symptoms appeared before the cutoff date. There were 60 patients in the CPT-11 + BSC arm and 62 patients in the BSC-only arm who were asymptomatic at baseline. For the CPT-11 + BSC arm and the BSC-only arm, the median symptom free survival was 8.1 months (95% CI: 6.1-10.7 months) and 7.0 months (95% CI: 4.4-8.7 months) respectively. The unadjusted logrank p-value was 0.2303.

**Time Until Worsening of Pain**

For the CPT-11 + BSC arm and the BSC-only arm, the median time to performance status deterioration was 10.3 months (95% CI: 7.8- . months) and 8.5 months (95% CI: 6.2-10.2 months) respectively. The unadjusted logrank p-value was 0.0586.

Among CPT-11 + BSC patients who did not take analgesics at baseline, 6.2% were taking opioids and 6.2% were taking non-opioids at week 18. Among BSC-only patients who did not take analgesics at baseline, 11.1% were taking opioids and 12.3% were taking non-opioids at week 18. For those patients who were not taking opioids at baseline, 6.7% of patients on the CPT-11 + BSC arm were taking opioids at 18 weeks, compared to 13.8% of patients on the BSC-only arm.

**Time to Performance Status Deterioration**

For the CPT-11 + BSC arm and the BSC-only arm, the median time to performance status deterioration was 6.4 months (95% CI: 5.2-7.6 months) and 5.1 months (95% CI: 4.2-6.9 months) respectively. The survival curves of this endpoint crossed at approximately 4.5 months. The unadjusted logrank p-value was 0.1865, although the interpretation of this p-value becomes difficult if the standard distributional assumptions do not hold.

**Time to Weight Loss Greater Than Five Percent**

For the CPT-11 + BSC arm and the BSC-only arm, the median time to performance status deterioration was 8.9 months (95% CI: 6.7-12.3 months) and 7.4 months (95% CI: 4.7-11.6 months) respectively. The unadjusted logrank p-value was 0.2279, which is statistically significant.

**Cox Analysis of Survival**

As a secondary analysis, a Cox model adjusted for a group of prognostic factors was fitted for survival. Some of the prognostic factors were specified in the protocol, while others were additionally added in order to take into account recent developments in the medical literature. The clinical prognostic factors appear in Figure 2. A Cox univariate analysis was performed to select the significant variables for a multiple Cox regression. Variables were selected for the multiple regression if they were statistically significant in a univariate model using an alpha of 0.10. The variables that were significant at 0.10 in their respective univariate models were performance status (0 vs $\geq$ 1), organ involvement, liver mets, tumor location (colon right vs. other), weight loss, 5FU bolus (Y/N), platelets, hemoglobin, white blood cells, alkaline phosphatase, SGOT, SGPT, and bilirubin.

In the multiple regression on these variables, a stepwise method was chosen, where the p-value to enter the model was 0.10 and the p-value to stay in the model was 0.05. The results of this model appear in Table 3.7.

**Table 3.7.** Results of stepwise Cox regression on survival using the prognostic factors in Figure 1. Treatment group was purposely withheld from this model.

| Prognostic Factor | Risk Ratio | p-value |
|---|---|---|
| WHO performance status (0 vs ≥ 1) | 1.538 | 0.012 |
| Tumor location (colon right vs. other) | 1.465 | 0.048 |
| SGOT (< 1.26 norm vs. ≥ 1.26 norm) | 1.533 | 0.034 |
| White Blood Cells (≥ 9 vs. <9 x$10^9$/l) | 1.620 | 0.009 |
| Hemoglobin (≥11.9 vs. 10.9-11.9 g/dl) | 1.832 | 0.005 |
| Hemoglobin (≥11.9 vs. 0.0-10.9 g/dl) | 2.120 | 0.003 |
| Alk. Phosphatase (0-126 vs. 126-250) | 1.572 | 0.026 |
| Alk. Phosphatase (0-126 vs. 250) | 2.537 | <0.001 |

When one adds the additional variable of treatment group into this model, the treatment group continues to be significant (risk ratio = 1.453; p-value = 0.017).

As in study , the sponsor classified numeric variables such as hemoglobin, alkaline phosphatase, and others into indicator variables. As before, it was not clear to this reviewer as to how the cutoff points for each subdivision were determined, although it appears as if they were chosen for nearly-equal distribution of patients across the selected number of intervals. See Table 3.8 for examples of the subdivisions of the indicator variables. Note that none of the cutoff points in Table 3.8 are consistent with the cutoff points used for the Cox regression (Table 3.3).

**Table 3.8.** Subdivisions of variables considered in the Cox models in V302. The units for each variable were not specified in the subdivisions.

| Platelets | | WBC | | SGPT | | SGOT | | Alk. Phos. | |
|---|---|---|---|---|---|---|---|---|---|
| Subdiv. | N | Subdiv. | N | Subdiv. | N | Subdiv. | N | Subdiv. | N |
| [0, 210) | 63 | [0, 6) | 64 | [0, 40) | 59 | [0, 52) | 62 | [0, 76) | 61 |
| [210, 251) | 63 | [6, 7) | 61 | [40, 60) | 66 | [52, 77) | 66 | [76, 105) | 61 |
| [251, 324) | 65 | [7, 9) | 70 | [60, 90) | 61 | [77, 126) | 68 | [105, 126) | 36 |
| [324, ∞) | 65 | [9, ∞) | 61 | [90, ∞) | 64 | [126, ∞) | 55 | [126, ∞) | 86 |

Although treatment arm was significant in the final Cox model for survival, it is this reviewer's opinion that the significance for white blood cells, SGOT, hemoglobin, and alkaline phosphatase in Table 3.7 should be interpreted with caution since their categorical analog may not be based on clinically relevant cutoff points.

This reviewer again performed an exploratory analysis for those variables with clear, clinically relevant subdivisions with assistance with the medical reviewer. Table 3.9 shows the covariates considered as well as the cutoffs used for a multiple Cox regression. The variable selection method for this regression analysis was the same as the sponsor's; namely, the variables that were significant at 0.10 in their respective univariate models were eligible for the final stepwise selection procedure. We also performed a backward elimination selection procedure (p-value of 0.05 required for retention) and a forward selection procedure (p-value of 0.10 required for entry) to investigate the consistency of our model-building.

The univariate regression p-values appear in the rightmost column Table 3.9. The variables that were eligible for the selection procedure were performance status, liver metastases, tumor location, and type of last prior chemotherapy, which is a different subset of variables from those selected by the sponsor.

**Table 3.9.** Variables and cutoff values for an exploratory Cox regression for survival in Study The "P-val." column refers to the p-value for the univariate Cox regression.

| Variable | Cutoffs | P-val. |
|---|---|---|
| Age | 0 to 64 vs. 65 and older | 0.456 |
| Gender | Male vs. female | 0.372 |
| Performance Status | 0 or 1 vs. 2 | 0.001 |
| Num. Of Organs Involved | 1 to 3 vs. more than 3 | 0.166 |
| Intent of prior chemo. | Adjuv. vs. advanced or adv. adjuv. | 0.641 |
| Best Resp. on 5FU | CR or PR vs. SD or PD | 0.531 |
| Liver metastases | Yes vs. no | 0.079 |
| Tumor location | colon right vs. other | 0.026 |
| Type of last prior chemo. | 5FU bolus vs. non-bolus | 0.026 |

The results of the variable selection procedures appear in Table 3.10.

**Table 3.10.** The results of three variable selection techniques using the variables in Table 3.9 in a Cox regression on survival. "Vars. selected" indicated the variables selected in the final model and "+Treatment" indicates the p-value when treatment arm is included in the final model.

| Stepwise | | Forward | | Backward | |
|---|---|---|---|---|---|
| Vars. selected | p-val | Vars. selected | p-val | Vars. selected | p-val |
| Last chemo. | 0.006 | Last chemo. | 0.008 | Last chemo. | 0.008 |
| Perf stat. | 0.001 | Perf stat. | 0.001 | Perf. stat. | 0.001 |
| | | Liver mets. | 0.042 | Liver mets. | 0.042 |
| +Treatment | 0.068 | +Treatment | 0.032 | +Treatment | 0.032 |

Note that the stepwise procedure selects a different model than the forward or backward procedures. Also, treatment arm is not significant in the final model selected by the stepwise procedure.

**4. Quality of Life**

To assess quality of life (QOL) analysis, the sponsor compared results of the EORTC QLQ-C30 instrument In the following discussion, the unit of time that we consider for will be weeks and the unit of time for will be visit, which will make the arms much more comparable. Figures 3 and 4 show the compliance to the QOL instrument for respectively. Compliance to the questionnaire completion in both studies were approximately 80% from baseline to week 12 and then gradually decreased over the follow up period. The sponsor used a logistic regression approach to test for differences in compliance between the CPT-11 arm and the control arm of both studies. There were no statistically significant differences with respect to questionnaire compliance in both V301 ($p = 0.38$) and V302 ($p = 0.46$).

**Figure 3**. Compliance to the QOL instrument in Study

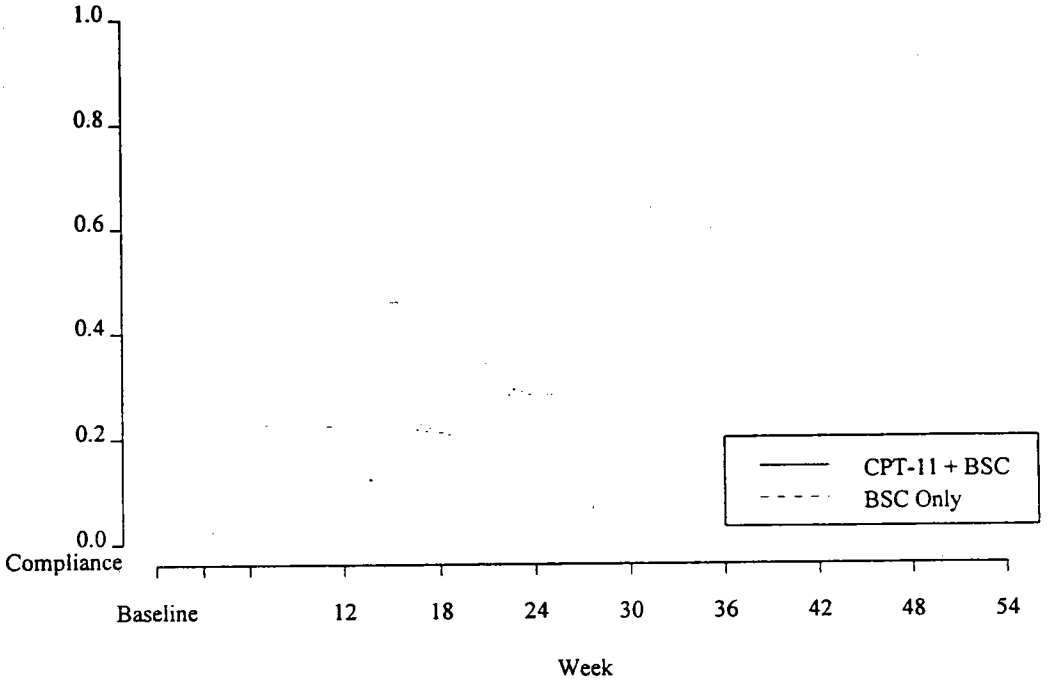
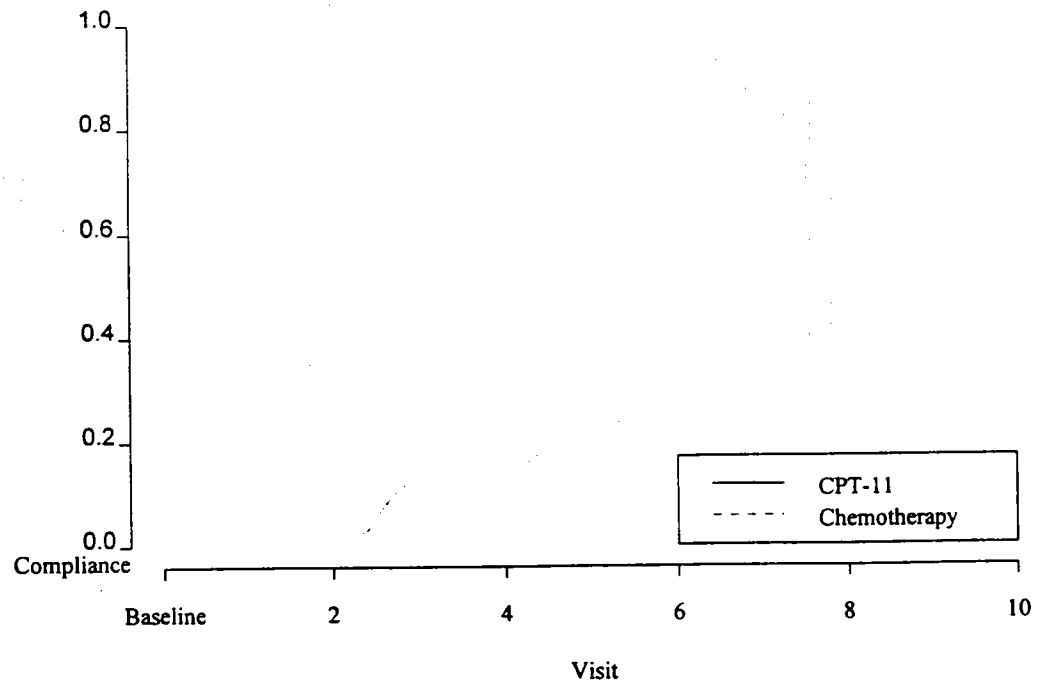



**Figure 4**. Compliance to the QOL instrument in Study



Using MANOVA and univariate techniques, the sponsor tested for differences in QOL at baseline. The MANOVA tests resulted in no global statistically significant difference in both V301 ($p = 0.11$) and V302 ($p = 0.97$). For V301, univariate tests on each subscale showed that statistically significant differences existed between treatment arms for Physical Functioning ($p = 0.0247$, in favor of CPT-11 + BSC), Fatigue ($p = 0.0091$, in favor of BSC-only) and Appetite Loss ($p < 0.0001$, in favor of BSC-only). Univariate tests on each subscale in resulted in p-values no smaller than 0.13.

For both studies, the sponsor first performed a global MANOVA for the instrument as a whole (all subscales simultaneously). For the global MANOVA analysis, treatment arm was significant in both studies ($p = 0.0001$ for and $p = 0.0002$ for

The sponsor fit a general linear model for each subscale using treatment arm, time and treatment*time as explanatory variables. For response variables, the sponsor considered all of the following: raw scores, differences from baseline, worst score, change from baseline of worst score, and scores with 0 (worst) imputed for patients who died. Table 4.1 and Table 4.2 show the significant subscales for respectively, for ANOVA on raw scores and on worst scores as the response variable.

**Table 4.1.** QOL subscales that were significant in the sponsor's analysis

| Change from baseline | P-val (arm favored)* | Worst score | P-val. (arm favored)* |
|---|---|---|---|
| Cognitive. Funct. | <0.001 (A) | Physical Funct. | <0.001 (A) |
| Global Health | 0.003 (A) | Role Funct. | 0.002 (A) |
| Pain | 0.008 (A) | Cognitive Funct. | 0.006 (A) |
| Dyspnea | 0.035 (A) | Social Funct. | 0.009 (A) |
| Appetite Loss | <0.001 (A) | Fatigue | 0.006 (A) |
| Financial Impact | <0.001 (A) | Pain | 0.001 (A) |
| Diarrhea | 0.017 (B) | Dyspnea | 0.029 (A) |
| | | Appetite Loss | <0.001 (A) |
| | | Constipation | 0.004 (A) |
| | | Diarrhea | <0.001 (B) |

*Arm A refers to the CPT-11 + BSC arm and Arm B refers to BSC-only.

**Table 4.2.** QOL subscales that were significant in the sponsor's analysis

| Change from baseline | P-val (arm favored)* | Worst score | P-val. (arm favored)* |
|---|---|---|---|
| Cognitive. Funct. | 0.001 (B) | Nausea/Vomiting | 0.007 (B) |
| Nausea/Vomiting | 0.011 (B) | Diarrhea | 0.030 (B) |
| Diarrhea | 0.009 (B) | Financial Impact | 0.045 (B) |

*Arm A refers to the CPT-11 arm and Arm B refers to chemotherapy.

As expected, Diarrhea was statistically significantly worse on the CPT-11 arms in both studies, and Nausea/Vomiting was significantly worse on the CPT-11 arm in study . There was substantial evidence for improved QOL in favor of CPT-11 in study

This reviewer had several concerns with respect to this analysis. The sponsor reported no method of controlling Type I error to account for the number of QOL subscales that were considered. Also, the sponsors' analyses assume that dropout is completely at random, which is not always the case in cancer trials. Analyzing QOL data presents us with two challenges. The first challenge is that of within-patient correlation across the repeatedly measured QOL endpoints. The second challenge is that of missing data.

In a classical univariate repeated ANOVA, a particular correlation structure known as compound symmetry must be assumed for a valid F-test of interaction of treatment and time. A multivariate approach may be considered when a compound symmetry assumption fails. However, in a multivariate approach, a distribution must be explicitly specified with the "correct" mean and covariance matrix.

The generalized estimating equation (GEE) approach was developed to cope with the potential problem of informative correlation among observations per subject. An advantage of a GEE approach is that it is not necessary to specify the correct correlation structure in advance. Using the idea of M-estimation theory (Huber, 1967; White, 1982; Liang and Zeger, 1986), the solution to the (potentially mis-specified) covariance matrix is consistent. Also, M-estimation protects the under-estimation of the covariance matrix by introducing "sandwich" estimators. Therefore, we have some assurance of a variance estimate that is robust.

This reviewer analyzed the QOL data using a GEE linear model and derived a robust covariance estimator based on M-estimation theory. To deal with the problem of potentially informative dropout, this reviewer based the dropout analyses on the concept of a pattern-mixture model (Little, 1993 and 1995).

Only three QOL subscales were considered in this reviewer's analysis. These were Physical Functioning, Pain, and Nausea/Vomiting. These were determined to be the most clinically relevant QOL measurements by both this reviewer and the medical reviewer.

It was determined from graphical methods and using complete and reduced modeling methods that, in the various treatment arms, the estimated linear trends for four out of the six subscales were noticeably different for those subjects who dropped out early as compared to those that completed later courses (see Figures 5-10). Therefore, this missing data pattern was concluded to be nonignorable for these four subscales and subjects were divided into two groups, those who completed no more than the third course and those who completed at least one course beyond the third course. For convenience, we will refer to the former group as "Dropouts" and the latter group as "Completers" throughout this analysis. The results of the QOL analyses appear in Table 4.3.

**Table 4.3.** QOL analysis of using longitudinal methods with complete and reduced models.

| QOL subscale | Subgroup | P-val | P-val |
|---|---|---|---|
| Physical Functioning | Dropouts | 0.58 | <0.001 |
| | Completers | 0.38 | <0.001 |
| Pain* | Dropouts | 0.031 | |
| | Completers | 0.28 | |
| Nausea/Vomiting* | Dropouts | 0.87 | |
| | Completers | 0.005 | |

* There was no evidence of a nonignorable missing data pattern for the Pain and Nausea/Vomiting subscales

**Figure 5.** Physical Functioning QOL subscale, Completers vs. Dropouts, . On this subscale, higher scores imply increased physical functioning. Completers are represented as the longer of the two sets of lines.

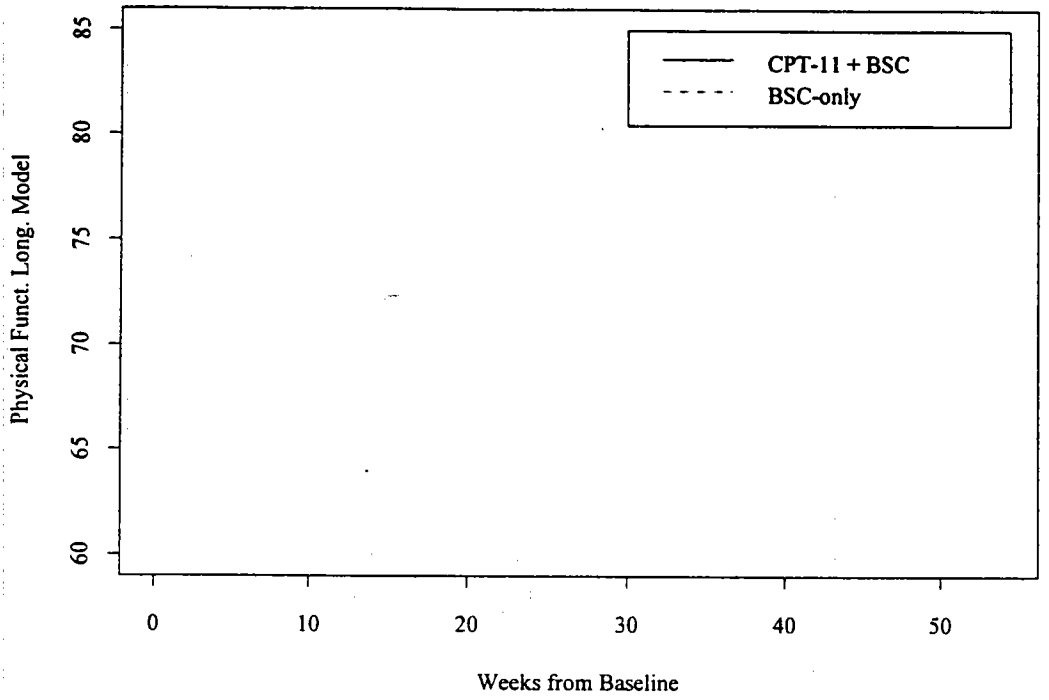



**Figure 6.** Pain QOL subscale, Completers vs. Dropouts, On this subscale, higher scores imply increased pain. Completers are represented as the longer of the two sets of lines.

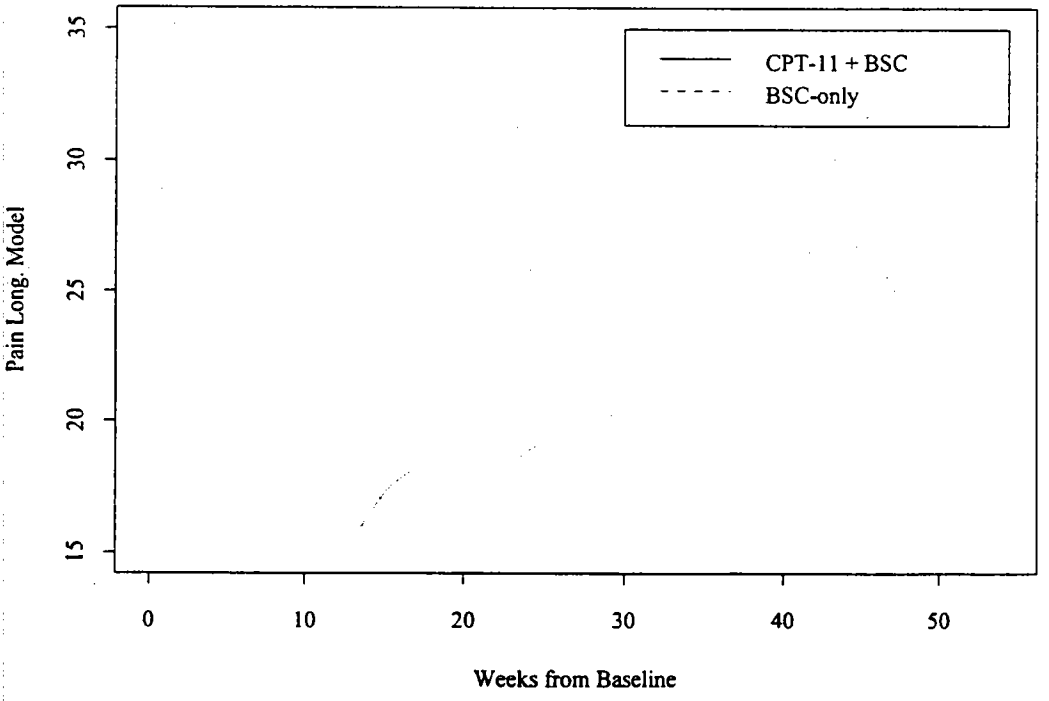


**Figure 7.** Nausea/Vomiting QOL subscale, Completers vs. Dropouts, On this subscale, higher scores imply increased nausea or vomiting. Completers are represented as the longer of the two sets of lines.

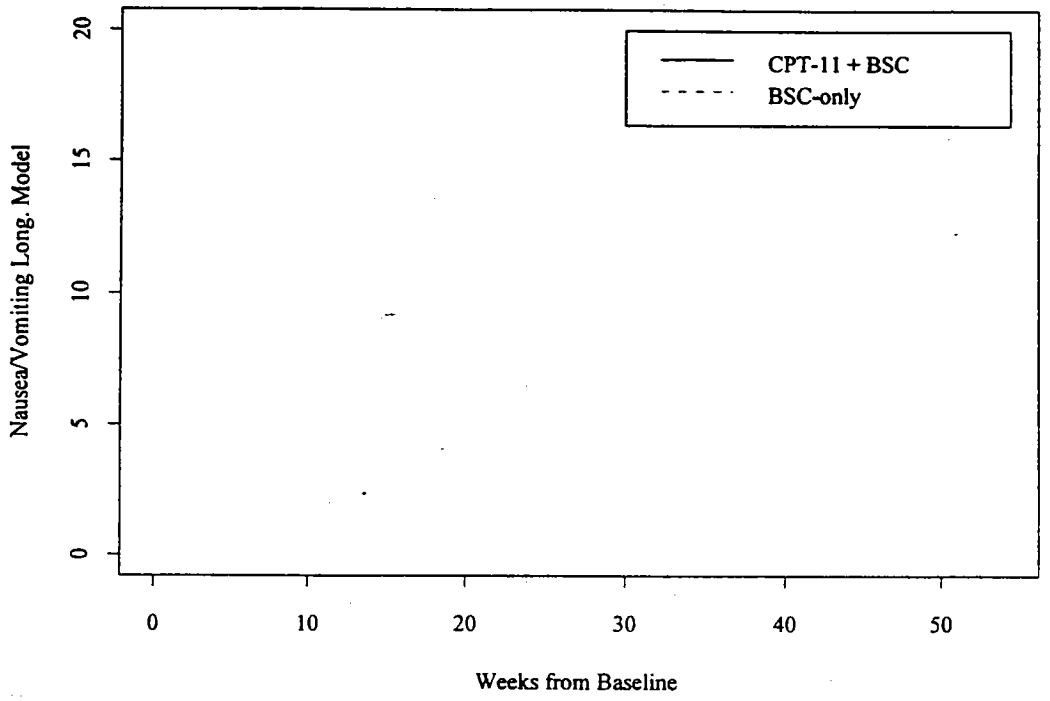



**Figure 8.** Physical Functioning QOL subscale, Completers vs. Dropouts, . On this subscale, high scores imply high physical functioning. Completers are represented as the longer of the two sets of lines.

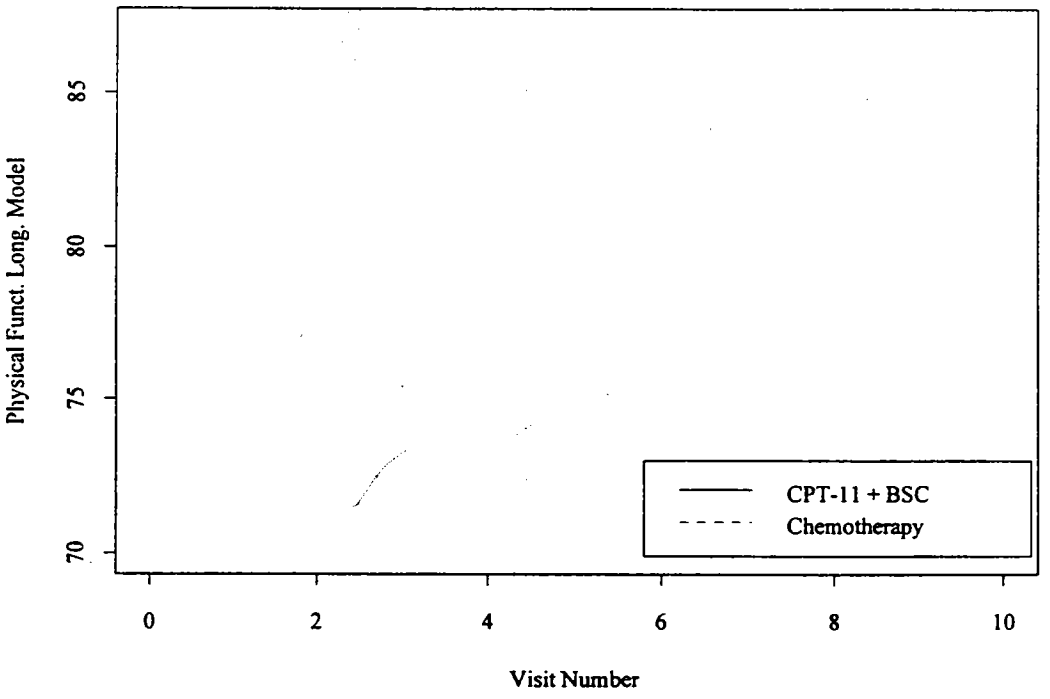


**Figure 9.** Pain QOL subscale, Completers vs. Dropouts, On this subscale, higher scores imply increased pain. Completers are represented as the longer of the two sets of lines.

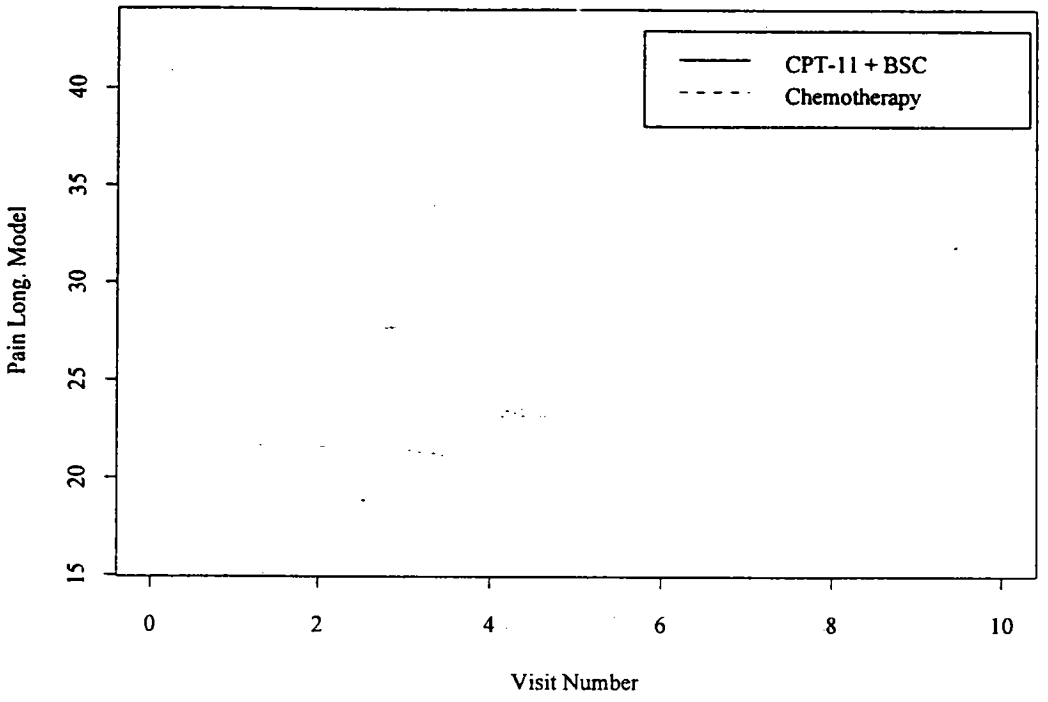
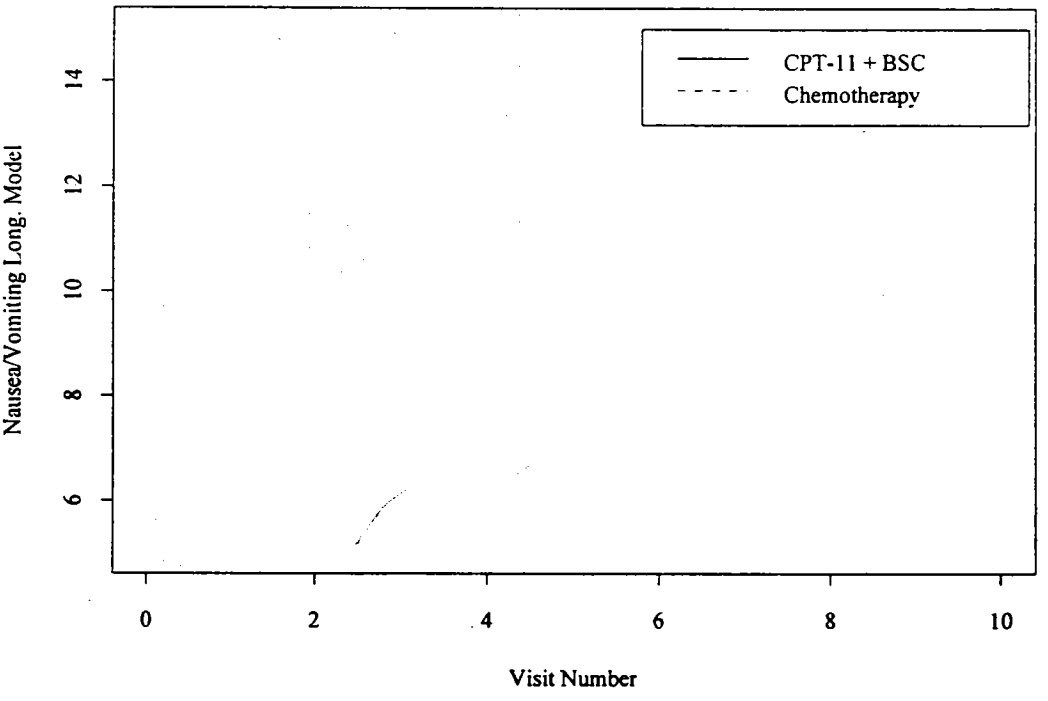



**Figure 10.** Nausea/Vomiting QOL subscale, Completers vs. Dropouts, On this subscale, higher scores imply increased nausea or vomiting. Completers are represented as the longer of the two sets of lines.



**QOL:**

In the Physical Functioning subscale, there was no statistically significant difference between the CPT-11 + BSC arm and the BSC-only arm either in Dropouts ($p = 0.58$) or Completers ($p = 0.38$). In the Pain subscale, there was no statistically significant difference between the CPT-11 + BSC arm and the BSC-only arm in Completers ($p = 0.28$) but Dropouts demonstrated a smaller increase in pain in the CPT-11 + BSC arm compared to the BSC-only arm ($p = 0.031$). Finally, in the Nausea/Vomiting subscale, there was no statistically significant difference between the CPT-11 + BSC arm and the BSC-only arm in Dropouts ($p = 0.87$), but Completers demonstrated a statistically significantly smaller increase in nausea/vomiting scores in the CPT-11 + BSC arm ($p = 0.005$).

**QOL:**

In the Physical Functioning subscale, there was a statistically significant difference in Physical Functioning responses for both Dropouts and Completers. In both cases, the CPT-11 + BSC arm deteriorated at a higher rate than the chemotherapy arm ($p < 0.001$ for both Dropouts and Completers). Using tests of complete and reduced models, this reviewer determined that there was little evidence of nonignorable dropout for the Pain and Nausea/Vomiting subscales. Therefore, one may conclude that the analyses performed by the sponsor for these subscales should not be biased from dropout to the extent that their results are invalid. According to the sponsor's analysis, there were no statistically significant differences on the Pain subscale, but significant differences on the Nausea/Vomiting subscale (see Table 4.2).

The results of this reviewer's QOL analysis is that the CPT-11 arms show statistically significant increase in Nausea/Vomiting symptoms, which may not be consistent with the known toxicity of CPT-11. The sponsor found statistically significant improvements on a wide range of QOL symptoms measured by the QLQ-C30, although this reviewer has concerns about the validity of the conclusions due to the problems of dropout and deaths. Some of the statistically significant findings of the sponsor were reproducible between whereas only Nausea/Vomiting and Pain differences were statistically significant in this reviewer's analysis. As it is particularly difficult to assess and analyze QOL in cancer trials, we must be careful not to rule out possible QOL benefits of CPT-11 in light of the fact that these two analyses draw differing conclusions. However, QOL should be considered as a supportive secondary endpoint with respect to the efficacy of CPT-11.

## 5. Summary and Conclusions

These studies were designed to test the superiority in the primary endpoint of survival of CPT-11 versus control arms of best supportive care and chemotherapy regimens. In the intent-to-treat population of the Phase III trial, median survival for the CPT-11 + BSC arm was 9.2 months, versus a median survival of 6.5 months for the BSC arm. This difference was statistically significant ($p < 0.001$). In the intent-to-treat population of the Phase III trial, median survival for the CPT-11 arm was 10.8 months, versus a median survival of 8.5 months for the chemotherapy arm. This difference was statistically significant ($p = 0.035$).

In the CPT-11 arm was superior on other secondary endpoints that were considered. For CPT-11 + BSC versus BSC-only, time until pain, time until performance status deterioration, and time until weight loss were all statistically significant at 0.05. These endpoints were considered in but they were not statistically significant.

Cox modeling was performed both by the sponsor and by this reviewer. For the Cox analyses in both studies, treatment arm was statistically significant.

This reviewer could not find the rationale for subdividing variables such as SGOT, SGPT, alk. phos., bilirubin, and WBC. The subdivisions were not the same between the two studies for some variables,

and it appears as though the subdivisions were based on equal sample size allocation instead of clinically relevant cutoffs. Therefore, there may be some difficulty in interpreting the statistical relevance of these variables.

The sponsor found QOL advantages of CPT-11 on many of the subscales that were considered, particularly in This reviewer had concerns about the effect that dropout and deaths had on the QOL conclusions and performed a separate analysis on only three selected subscales. The sponsor's analysis and this reviewer's analysis did not agree on the fact that patients on the CPT-11 arms had increased nausea and vomiting as compared to the patients on the control arms; this reviewer concluded that patients on CPT-11 less nausea and vomiting compared with the control arm while the sponsor concluded that patients on CPT-11 had more nausea and vomiting compared to the control arm. The sponsor's analysis is more consistent with the known increased toxicity of CPT-11. The other subscales that were considered were not consistent between the two separate analyses or between the two studies. This reviewer had difficulty supporting the claim that there is evidence of QOL improvements in general in patients on CPT-11.

**6. Overall Recommendations and Conclusions**

In the two Phase III trials included in this submission, survival was the primary endpoint. There is substantial evidence to conclude that CPT-11 prolongs survival in patients with colorectal cancer. The Cox regression analyses provided futher supportive evidence that CPT-11 is superior to either control arms in terms of survival.

There is evidence to conclude that CPT-11 also prolongs time until pain, time until weight loss, and time until performance status worsening. These endpoints were statistically significant in but were not statistically significant in

It is this reviewer's opinion that CPT-11 has demonstrated efficacy for the proposed indication based on the well-designed and well-analyzed pivotal trials that were submitted.

/S/

David Smith, Ph.D.
Mathematical Statistician

**References**

Huber, P.J. "The behavior of maximum likelihood estimates under nonstandard conditions," in *Proceedings of the fifth Berkeley Symposium in Mathematical Statistics and Probability*, Berkeley: University of California Press, 221-233, 1967.

Little, R.J.A. "Pattern-mixture models for multivariate incomplete data," *JASA*, **88:** 125-134, 1993.

Little, R.J.A. "Modeling the drop-out mechanism in repeated-measures studies," *JASA*, **90:** 1112-1121, 1995.

Liang, K.Y. and Zeger, S.L. "Longitudinal data analysis using generalized linear models," *Biometrika*, **73:** 13-22, 1986.

White, H. "Maximum likelihood estimation of misspecified models," *Econometrica*, **50:** 1-25, 1982.

Concur:

Dr. Chen /S/

Dr. Chi /S/ 8/3/98

cc: Archival NDA 20-571

HFD-150 / Mr. P. Guinn, Project Manager

HFD-150 / Dr. I. Chico

HFD-150 / Dr. G. Williams

HFD-344 / Dr. B. Barton

HFD-710 / Dr. G. Chi

HFD-710 / Dr. G. Chen

HFD-710 / Dr. D. Smith

HFD-710 / Chron

Smith / 3 August 1998 / Word / c:\data\wordfiles\20571.doc

This review consists of 21 pages of text.