Project Assignment

Each year cancer of the colon will afflict over 100,000 Americans. Eighty percent present at a sufficiently early stage that all clinically diagnosed disease can be surgically removed. Those which have regional nodal involvement which is clinically completely resected are referred to as having Duke's Stage C disease. Unfortunately, one half of these patients will have residual cancer existing in an occult and probably microscopic stage which will lead to recurrence of disease and death within 5 years. Although numerous randomized trials of adjuvant therapy for resected colon cancer had been conducted in the past, involving thousands of patients, there had been no convincing and reproducible evidence that either chemotherapy or immunotherapy produce sufficient benefit to justify their risks or toxicity. Until very recently, the consensus had been that the best standard therapy was surgery alone. Recent reports (*J. Clin. Oncology* 1989;71447-1456 and *New England J. of Medicine* 1990;322: 352-358) suggest benefit from the use of levamisole in combination with 5-fluorouracil (5-FU), as adjuvant therapy for Stage C colorectal cancer.

You have been provided with the data on the Duke's C patients who were reported in the *J. Clin. Oncology* and in the *New England Journal of Medicine* article. The latter article provided median follow-up of 3.5 years, which was available in late September 1989 when the Data Monitoring Committee judged results to be definitive. Your data set has considerably richer long term information, with approximately 6.5 years median follow-up on patients.

Assignment

a) Provide a detailed analysis to address the influence of treatment on time to disease recurrence and on patient survival.

b) These data provide a large and rich resource for studying the natural history of this disease. Provide an in depth analysis to build a parsimonious model which can be used to predict survival of future patients.

In your report, discuss (or better yet carry out) methods for model validation.

The report of your analyses should be in a form readable both by statisticians and by clinicians.

BACKGROUND FOR ANALYSIS OF ADJUVANT COLON CANCER DATA

Colon cancer typically starts on the lining of the inside of the bowel wall (mucosa). Next it grows into the bowel wall (muscular involvement) and then proceeds through and bulges on the outside of the wall into the outer lining (serosa). It then might proceed further to adhere to adjacent organs (e.g., stomach, pancreas, or abdominal wall) and ultimately to invade those adjacent organs. This explains EXTENT OF LOCAL SPREAD and SEROSAL INVOLVEMENT.

ADHERENCE indicates whether or not the tumor has grown to adhere to an adjacent organ (it can be "peeled away" since it has not invaded).

ORGAN INVOLVEMENT indicates whether the tumor actually has invaded an adjacent organ.

PERFORATION indicates whether the tumor has eroded a hole in the wall of the bowel by growing through and then becoming necrotic. Infection often follows. Some think this could be beneficial through stimulation of the immune system.

REGIONAL IMPLANTS exist when many tumor implants are found in the abdominal lining in the local region of the tumor, having been "seeded" like a dandelion deed.

OBSTRUCTION indicates whether the surgeon or pathologist has noted evidence of obstruction by the lesion, i.e., whether the bowel has become blocked, prohibiting free flow of fecal material, thought by some to be an indication of more advanced malignant disease.

The tumor's growth often leads it to an invasion of the lymph system, allowing seeds to spread through lymph channels. NODAL INVOLVEMENT indicates the number of resected lymph nodes involved with cancer.

If the tumor invades veins, seeds can be spread by the blood stream throughout the body. This often leads to metastatic involvement of critical organs such as the liver and lung.

HISTOLOGIC TYPE characterizes the type of colon cancer cells which are growing.

DIFFERENTIATION relates to the degree of development of the cancer cells. Well-differentiated cells are better developed, closer to normal healthy cells, and less aggressive or out of control in growth characteristics.

DAYS POST-OPERATION indicated the time between surgical excision and randomization to adjuvant therapy.

The only patients eligible for this study were those whose disease was "clinically" completely removed through an "en bloc" resection having tumor free margins.

DOCUMENTATION FOR COLON CANCER DATASET

THe description of the data is as follows

Variable Name(*)	Variable description	Code	Code Definition
PATNO	Patient number		
TRNO	Treatment	1	Observation
		2	Levamisole alone
		3	5-FU + Levamisole
DAYREG	Day of registration		
MTHREG	Month of registration		
YRREG	Year of registration		
DAYLC	Day of last contact or death		
MTHLC	Month of last contact or death		
YRLC	Year of last contact or death		
STATUS	Survival status	0	Alive
		1	Dead
DAYPRGREL	Day of prog/relapse		
MTHPRGREL	Month of prog/relapse		
YRPRGREL	Year of prog/relapse		
STRATNO	Stratification level (see note @ below)		
SEX	Sex	0	Female
		1	Male
PT1SITE	Location of primary neoplasm	1	Cecum
		2	Right Colon
		3	Hepatic flexure
		4	Transverse colon
		5	Splenic flexure
		6	Left colon
		7	Sigmoid colon
		8	Rectosigmoid
		9	Rectum
		10	Multiple sites
HISTYPE	Histologic type	1	Adenocarcinoma
		2	Colloid(mucinous)
		3	Signet ring type
		4	Other
CDIFINTG	Differentiation	1	Well
		2	Moderate well (gr 2-3)
		3	Poor (grade 4)
EXTLOC	Extent of local spread	1	Submucosa/not muscle
		2	Muscular/not serosa
		3	Serosa/not contiguos
		4	Contiguos structures

OBSTRC	Obstruction	0	No
		1	Yes
PERFOR	Perforation	0	No
		1	Yes
ADHERE	Adherence	0	No
		1	Yes
REGIMP	Regional Implants	0	No
		1	Yes
CEAPREOP	Pre-operative CEA level, missing for many patients		
NOPOSND	Nodal involvement (number of positive nodes)		
DAYRES	Day of tumor resection		
MTHRES	Month of tumor resection		
YRRES	Year of tumor resection		
DAYRX	Day of start of treatment		
MTHRX	Month of start of treatment		
YRRX	Year of start of treatment		
PRGREL	Prog/relapse status	0	No
		1	Yes
AGE	Patient age at registration		
DAYS	Number of days from tumor resection to start of treatment.		
TIME	Time for registration to death or last contact (see status)		
GROUP	1: Data from SWOG 8591(aka Int - 0035) 2: Data from NCCTG		

(*) You can name the variables whatever you want. Most of the names are from the original SAS dataset.

@STRATNO -- possible values range from 1-24

If stratno is in (2,5,8,11,14,17,20,23) then lymph nodes = 1-4, if stratno is in (3,6,9,12,15,18,21,24) then lymph nodes > 4. If stratno is in (1,4,7,10,13,16,19,22) then lymph nodes = 0(by definition of dukes B).

If stratno is in 1-12 then days from surgery to registration = "7-20 days", otherwise if stratno is in 13-24 then days from surgery to registration = "21-30 days".

This additional information is useful because some patients have a missing value for "number of positive lymph nodes" or "resection date". By using a patient's stratification number, one is able to categorize the patient without knowing the exact value.