Documentation for crypt.txt

Colon cancer is one of the leading causes of cancer morbidity and mortality in the United States. Increasing attention has been focused on prevention of this disease. Basic preventive strategies have focused on diet modification, existing pharmaceutical agents such as nonsteroidal anti-inflammatory drugs (which have been approved for public use for other indications, but not as a colon cancer preventive agent), and various experimental chemicals.

The interest in diet as a preventive agent for colon cancer arises largely from observational studies in which it has been observed that countries having larger fiber in their diet tend to have lower rates of colon cancer. Similarly, observational studies have suggested that persons with higher levels of calcium in their diet have lower incidence of colon cancer.

In an attempt to further establish these associations, a clinical trial was conducted to test whether dietary supplementation with fiber and/or calcium would lead to lower rates of cell division. The choice of this particular endpoint is based on the idea that cancer is a disease characterized by uncontrolled proliferation of cells. One theory is that persons at higher risk for cancer would first have higher rates of cell division among cells that otherwise appear normal.

In the clinical trial, patients who were thought to be at higher risk for development of colon cancer were equally likely to be randomized to any one of four groups:

1. low fiber, low calcium
2. low fiber, high calcium
3. high fiber, low calcium
4. high fiber, high calcium

Both the participants and the investigators were blinded to treatment assignment. Duration of treatment was 1 year, with cell division assessed at randomization, after 6 months of treatment, and after 1 year of treatment.

Cell division rates were determined using a method called thymidine crypt labeling. In this method, samples (biopsies) of the colon are obtained from each patient, and these biopsies are incubated with radioactive labeled thymidine. Because thymidine is a constituent of DNA, cells that are in the process of division will incorporate the radioactive thymidine into their chromosomes. After further processing, the cells with radioactive chromosomes can be identified under a microscope.

When the biopsies are examined microscopically, of particular interest are the cells in the colon crypts, which are fingerlike indentations in the colon mucosa. Usually, all cell division occurs toward the bottom of these crypts. In neoplastic diseases (such as cancer) it is thought that the increased growth of cells can be identified by an increased number and changing pattern of labeled cells. Specifically, there is some evidence in the literature that suggests that people prone to cancer have higher numbers of labeled (dividing) cells, and/or that those labeled cells occur at higher levels in the colon crypt than would happen in non-diseased persons.

The number of crypts that were of sufficient quality to allow counting of labeled cells might vary from biopsy to biopsy, and the number of cells in a given crypt also varied from crypt to crypt.

The question to be addressed in this analysis is whether fiber and/or calcium supplementation has affected the patterns of cell division in the colon crypts for the patients on study.

The data file crypt.txt contains the following information. Each line contains the data for a single crypt. The data on each line contains:

ptid	a patient identifier
seq	a variable identifying the sequence of biopsy (three biopsies were obtained for each patient: 1= randomization, 2= 6 months, 3= 1 year)
cntr	an indicator of the lab technician
tx	an indicator of treatment group as indicated above
left	an indicator of whether the count was the left or right side of the crypt
tcells	the total number of cells on that side of the crypt
nlbl	the number of labeled cells
l1, l2, ...	the position of the labeled cells (one number for each labeled cell): 1 is at the bottom of the crypt and cell nlbl is at the top

Note that the strange format of the file may be difficult to manage. The following S+ code can be used to read the file. It produces a list of the first 7 columns of the data file (named 'dta') and a matrix (named 'lblcells') of the labelled cell numbers with NAs for unused spaces in the matrix.

```
dta <- scan("crypt.txt",list(ptid=0,seq=0,cntr=0,tx=0,
             left=0,tcells=0,nlbl=0),flush=T)
lbl <- scan("crypt.txt",0)
maxlbl <- max(dta$nlbl)
lblcells <- matrix(NA,length(dta$ptid),maxlbl)
for (i in 1:length(dta$ptid)) {
             lbl <- lbl[-(1:7)]
             nlbl <- dta$nlbl[i]
             if (nlbl > 0) {
                          lblcells[i,1:nlbl] <- lbl[1:nlbl]
                          lbl <- lbl[-(1:nlbl)]
             }
}
```