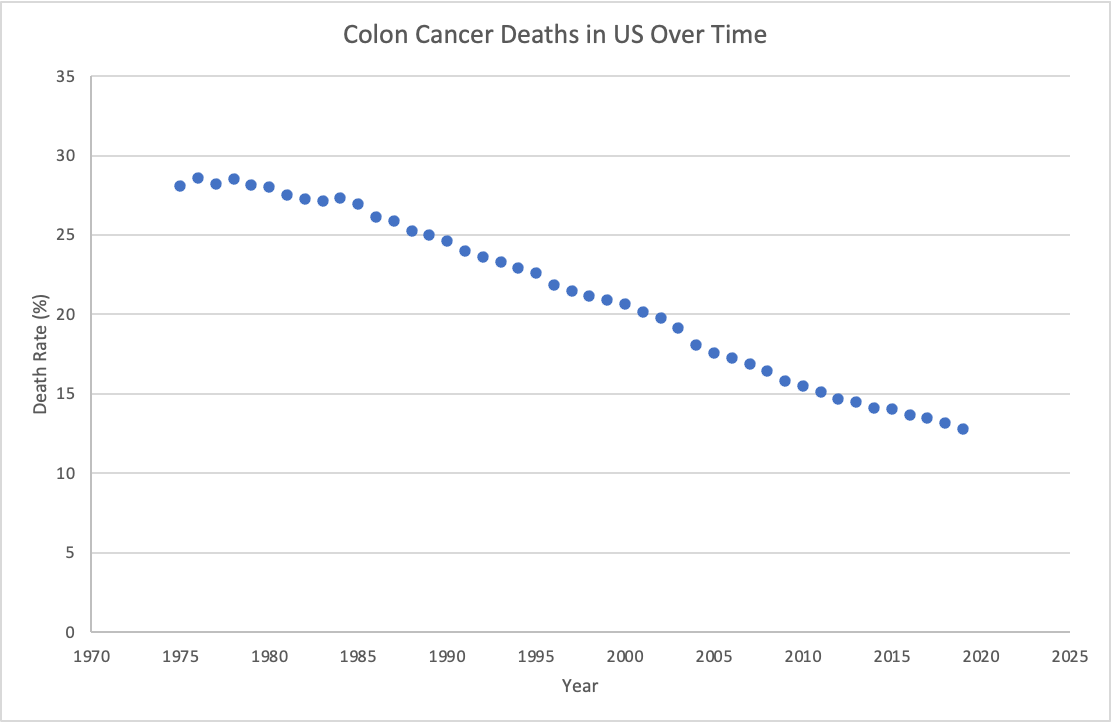
Introduction

Before the development of effective chemotherapy for colon cancer, treatment was limited to surgery. Recurrence was frequent and life-expectancy after initial treatment not long.

However, from 1985 to 1990, adjuvant chemotherapy began to be developed and implemented in combination with surgery.

After simultaneous advancements in detection and screening, death rates due to colon cancer began to fall.

Today, excluding skin cancer, colon cancer is the third most common type of cancer in the U.S., but there are now more treatments available, including immunotherapy and radiation therapy.



(GRAPH / DATA IN EXCEL FILE)

* <https://seer.cancer.gov/statfacts/html/colorect.html>
* <https://www.cancer.org/cancer/colon-rectal-cancer/about/key-statistics.html>

Data Source & Description of Variables

This dataset comes from one of the first successful trials of adjuvant chemotherapy for colon cancer, using a combination of levamisole and fluorouracil.

The trial examined both Stage B2 (locally invasive) and Stage C (regional nodal involvement) colon cancer, but this dataset contains information only on subjects with Stage C.

The dataset had a complicated format. For each subject, there were two rows of data entry: one measuring data related to the first recurrence of cancer, such as days until recurrence, and the other measuring data related to death, such as presence of cancer in other regions. The rows were categorized by the **etype** nominal variable, which we used to sort the data.

In addition, each row had a **status** nominal variable, that indicated whether the **etype** was positive or negative – whether there was or was not recurrence, and whether or not the subject had died. This required additional sorting of the data.

As each subject in the study was assigned a treatment – observation, levamisole (alone), or levamisole and fluorouracil – this was recorded by the **rx** nominal variable for each row of data, and allowed us to sort, analyze, and compare the effectiveness of the different treatments.

For the analysis, we looked at the discrete quantitative **time** variable (measured in days), and the nominal variable **extent**, which recorded whether the cancer had spread to specific nearby regions.

Research Questions

We sought to answer questions about the effectiveness of the different treatment types:

* Which treatment had a lower recurrence rate and by how much?
* Which treatment had a greater amount of time until recurrence and by how much?
* Which treatment had a higher survival rate and by how much?
* Which treatment had a higher life expectancy and by how much?