## Introduction to study

A cancer clinic wants to understand how four antineoplastic (e.g., anti-cancer) drugs are being given. Drugs A and B are chemotherapy drugs (sometimes given in combination) and Drugs C and D are immunotherapy drugs. The clinic has provided us with two datasets: one gives diagnoses by patient and the other dataset gives treatment dates for these patients for the drugs of interest. None of the patients in this cohort have died to date, and no data is missing.

## General questions

Based on the information provided above and the attached dataset, what three questions would you like to understand prior to conducting any analysis of the data?

The following questions I would like to understand prior to conducting data analysis.

1. What does the data represent and where does the data come from?

I would like to know the sourse of the data as it is crucial for assessing its reliability and relevance. What story does it telling which will include identifying its origin, collection methods, time-scale and any potential biases. I will like to know what all are the columns in the dataframe and what information they are storing.

1. Who are my stakeholders and what questions to they have?

I would like to find out who are my stakeholders and their expectations from me.

This will be helpful in aligning the analysis with stackholders needs and will make sure the outcomes are meaningful.

1. What is the goal of the study and what output we are looking for?

Understanding the goal of the study and the outcome from the analysis will provide me with clarity on desired output, assumptions if necessary, steps of analysis, study visualizations, and effective decision making.

## Data analysis questions

## Setting up my environment

Note: Setting up my environment by adding necessary packages including tidyverse, sarvival,etc.

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(survival)

## Uploading required dataframes

Note: Reading the Patient\_Diagnosis.csv and Patient\_Treatment.csv files.

Patient\_Diagnosis <- read.csv("Patient\_Diagnosis.csv")  
Patient\_Treatment <- read.csv("Patient\_Treatment.csv")

## Exploring Datasets

Note: I’ll check columns and column formats for consistency.

str(Patient\_Diagnosis)

## 'data.frame': 57 obs. of 4 variables:  
## $ patient\_id : int 2120 2720 2038 2238 2175 2475 2407 2607 2425 3025 ...  
## $ diagnosis\_date: chr "1/9/10" "1/9/10" "1/21/10" "1/21/10" ...  
## $ diagnosis\_code: num 174 174 175 175 175 ...  
## $ diagnosis : chr "Breast Cancer" "Breast Cancer" "Breast Cancer" "Breast Cancer" ...

str(Patient\_Treatment)

## 'data.frame': 1096 obs. of 3 variables:  
## $ patient\_id : int 2720 2238 2120 2038 2120 2038 2120 2120 2038 2038 ...  
## $ treatment\_date: chr "1/20/10" "1/21/10" "1/23/10" "1/24/10" ...  
## $ drug\_code : chr "B" "B" "B" "A" ...

head(Patient\_Diagnosis)

## patient\_id diagnosis\_date diagnosis\_code diagnosis  
## 1 2120 1/9/10 174.1 Breast Cancer  
## 2 2720 1/9/10 174.1 Breast Cancer  
## 3 2038 1/21/10 174.9 Breast Cancer  
## 4 2238 1/21/10 174.9 Breast Cancer  
## 5 2175 2/17/10 174.7 Breast Cancer  
## 6 2475 2/17/10 174.7 Breast Cancer

head(Patient\_Treatment)

## patient\_id treatment\_date drug\_code  
## 1 2720 1/20/10 B  
## 2 2238 1/21/10 B  
## 3 2120 1/23/10 B  
## 4 2038 1/24/10 A  
## 5 2120 1/24/10 A  
## 6 2038 1/24/10 B

summary(Patient\_Diagnosis)

## patient\_id diagnosis\_date diagnosis\_code diagnosis   
## Min. :2038 Length:57 Min. :153.3 Length:57   
## 1st Qu.:2770 Class :character 1st Qu.:153.8 Class :character   
## Median :4256 Mode :character Median :174.5 Mode :character   
## Mean :4899 Mean :168.0   
## 3rd Qu.:6889 3rd Qu.:174.9   
## Max. :9489 Max. :174.9

summary(Patient\_Treatment)

## patient\_id treatment\_date drug\_code   
## Min. :2038 Length:1096 Length:1096   
## 1st Qu.:2763 Class :character Class :character   
## Median :4976 Mode :character Mode :character   
## Mean :5055   
## 3rd Qu.:6840   
## Max. :9489

* str() returns columns, its datatypes and initial observations in dataframes.
* head() function will give details of first 6 rows.
* summary() will give the statistical summary as shown above. We got the basic details of both the dataframes.

**Explanation for exploring data:** Here, Patient\_id is in integer form. As we are not going to perform numerical operations on Patient\_id, we can convert it to character. diagnosis\_date and treatment\_date are in char format which ideally should be in date format. Other column formats also we will update.

## Updating Datatypes

Note: Updating datatypes will ensure all the variables in the specific column have specific format.

Patient\_Diagnosis <-  
 mutate( Patient\_Diagnosis,  
 patient\_id = as.character(patient\_id), #changing the datatype to character   
 diagnosis\_date = mdy(diagnosis\_date), #changing the datatype to mdy  
 diagnosis\_code = as.character(diagnosis\_code) #changing the datatype to character  
 )  
Patient\_Treatment <-  
 mutate( Patient\_Treatment,  
 patient\_id = as.character(patient\_id), #changing the datatype to character  
 treatment\_date = mdy(treatment\_date) #changing the datatype to mdy  
 )

## Explanation for updating datatypes: I changed the datatype of patient\_id, diagnosis\_date treatment\_date to character, mdy and mdy format resp. in Patient\_Diagnosis and Patient\_Treatment dataframes.

## Removing Duplicates

Note: We need to check for the repeated entries with same id and diagnosis. Multiple diagnosis is fine but multiple diagnosis of same patient type for same diagnosis should be neglected.

duplicate\_combinations <- Patient\_Diagnosis %>%  
 group\_by(patient\_id, diagnosis) %>%  
 summarise(count = n()) %>% #Calculate the number of occurrences (n(), which counts rows) for each combination of patient\_id and diagnosis  
 filter(count > 1) #Filter the grouped data to include combinations (patient\_id, diagnosis) where the count > 1, indicating duplicates

## `summarise()` has grouped output by 'patient\_id'. You can override using the  
## `.groups` argument.

print(duplicate\_combinations)

## # A tibble: 5 × 3  
## # Groups: patient\_id [5]  
## patient\_id diagnosis count  
## <chr> <chr> <int>  
## 1 3449 Colon Cancer 2  
## 2 3757 Colon Cancer 2  
## 3 4256 Breast Cancer 2  
## 4 4354 Breast Cancer 2  
## 5 4374 Breast Cancer 2

We have total 5 rows repeated.

diagnosis\_filtered <- Patient\_Diagnosis %>%  
 group\_by(patient\_id, diagnosis) %>%  
 slice\_min(diagnosis\_date, with\_ties = FALSE) %>%  
 ungroup()

slice\_min(diagnosis\_date, with\_ties = FALSE): Within each group (patient\_id and diagnosis), selects the row with the minimum diagnosis\_date. The argument with\_ties = FALSE ensures that if there are ties (multiple rows with the same minimum diagnosis\_date), only the first occurrence is kept. ungroup() will remove grouping.

Treatment\_filtered <- Patient\_Treatment %>% distinct()

This will remove the duplicates from Patient\_Treatment dataframe.

**Explanation for removing duplicates:** I created the dataset duplicate\_combinations which will include repeated values based on patient\_id and diagnosis.

Summarise (count=n()) calculates the number of occurrences (n(), which counts rows) for each combination of patient\_id and diagnosis.

filter(count > 1) will filter the grouped data to include combinations (patient\_id, diagnosis) where the count > 1, indicating duplicates.

5 such duplicate entries were found in Patient\_Diagnosis dataframe. I removed those entries from Patient\_Diagnosis dataset using slice\_min() function. slice\_min(diagnosis\_date, with\_ties = FALSE): Within each group (patient\_id and diagnosis), selects the row with the minimum diagnosis\_date. The argument with\_ties = FALSE ensures that if there are ties (multiple rows with the same minimum diagnosis\_date), only the first occurrence is kept. ungroup() will remove grouping.

To remove the duplicate entries from Patient\_Treatment, I used distint() function. Distinct() selects unique rows from a dataframe. It removes duplicate rows based on all columns by default.

### Question 1

the clinic would like to know the distribution of cancer types across their patients. Please provide the clinic with this information.

Ans: To understand the distribution of cancer type across the patients, we will group the data yearwise. First we will extract the number of patients for each cancer type. We will get the number patients having colon cancer and breast cancer. After that we will calculate the yearwise numbers. We will join the dataframes. Finally, we will represent the data graphically.

**Steps:**

Note: We will represent the yearwise data for each cancer type.

all\_dx <- diagnosis\_filtered %>%  
 group\_by(diagnosis) %>% #group data by diagnosis  
 count() %>% #calculate frequency of each unique diagnosis store it as n  
 pivot\_wider(names\_from = diagnosis, values\_from = n) %>% #reshape data in wider form  
 mutate("Diagnosis Year" = "All", .before = everything()) #.before = everything() argument ensures this new column is added at the beginning  
  
dx\_by\_year <- diagnosis\_filtered %>%  
 group\_by(diagnosis, year(diagnosis\_date)) %>%   
 count() %>% #calculate frequency of each unique combination of diagnosis and year  
 pivot\_wider(names\_from = diagnosis, values\_from = n) %>% #reshape data in wider form  
 rename(`Diagnosis Year` = `year(diagnosis\_date)`) %>%   
 mutate(  
 `Colon Cancer` = ifelse(is.na(`Colon Cancer`), 0, `Colon Cancer`), #Check if there are missing values (NA) in the Colon Cancer column and replaces them with 0 if true.  
 `Diagnosis Year` = as.character(`Diagnosis Year`)  
 )

pivot\_wider() transforms the data from long to wide format, for each cancer type we will check the total number of patients. Colon Cancer is replaced with 0 where it is NA using ifelse(is.na()).

#### creating template

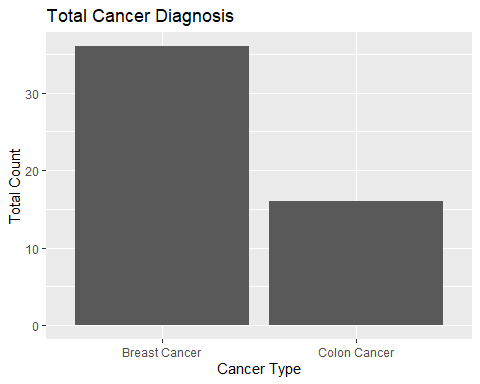
print(diagnosis\_gt <-  
 bind\_rows(  
 all\_dx,  
 dx\_by\_year)) #join the two tables

## # A tibble: 5 × 3  
## `Diagnosis Year` `Breast Cancer` `Colon Cancer`  
## <chr> <int> <dbl>  
## 1 All 36 16  
## 2 2010 10 0  
## 3 2011 7 8  
## 4 2012 10 6  
## 5 2013 9 2

This output shows that there are total 36 breast cancer patients and 16 colon cancer patients. Year 2010 and 2012 showed max breast cancer patients and Year 2011 showed max colon cancer patients.

To represent the data graphically, I used ggplot(), first graph is showing the cancer type vs total count. Second graph shows the yearwise distribution of cancer patients and diagnosis.

all\_dx\_graph <- diagnosis\_filtered %>%  
 ggplot(aes(x = diagnosis)) +  
 geom\_bar() +  
 labs(  
 x = "Cancer Type",  
 y = "Total Count",  
 title = "Total Cancer Diagnosis"  
 )  
  
print(all\_dx\_graph)



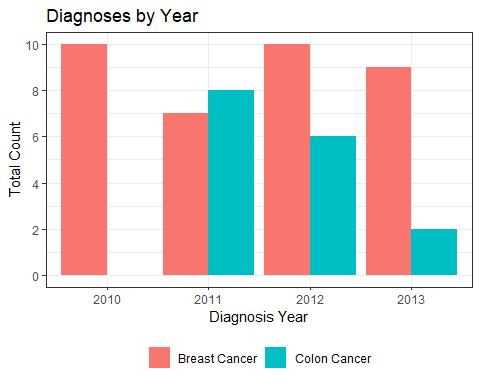
library(dplyr)  
library(ggplot2)  
library(scales) # For pretty\_breaks function

##   
## Attaching package: 'scales'

## The following object is masked from 'package:purrr':  
##   
## discard

## The following object is masked from 'package:readr':  
##   
## col\_factor

dx\_by\_year\_graph <- diagnosis\_filtered %>%  
 mutate(diagnosis\_year = lubridate::year(diagnosis\_date)) %>%  
 group\_by(diagnosis\_year, diagnosis) %>%  
 summarise(count = n(), .groups = 'drop') %>% # Calculates the count of each unique combination of diagnosis\_year and diagnosis, .groups='drop' acts as ungroup  
 complete(diagnosis\_year = min(diagnosis\_year):max(diagnosis\_year),  
 diagnosis,  
 fill = list(count = 0)) %>% #Ensures that the data is complete for all combinations of diagnosis\_year and diagnosis within the range from the minimum to the maximum diagnosis\_year. Missing combinations are filled with count = 0  
 ggplot(aes(x = factor(diagnosis\_year), y = count, fill = diagnosis)) +  
 geom\_bar(stat = "identity", position = "dodge") +  
 labs(  
 x = "Diagnosis Year",  
 y = "Total Count",  
 title = "Diagnoses by Year",  
 fill = "Diagnosis"  
 ) +  
 theme\_bw() +  
 scale\_y\_continuous(breaks = scales::pretty\_breaks()) +  
 theme(  
 legend.title = element\_blank(),  
 legend.position = "bottom"  
 ) #Sets breaks for the y-axis using pretty\_breaks() for visually appealing breaks  
  
print(dx\_by\_year\_graph)



### 

### Question 2

The clinic wants to know how long it takes for patients to start therapy after being diagnosed, which they consider to be helpful in understanding the quality of care for the patient. How long after being diagnosed do patients start treatment?

Ans: Here, we need to find out the gap between diagnosis date and the treatment date.

For that, I first merged both the dataframes (diagnosis\_filtered and Treatment\_filtered). I thought if I sort the patient\_id and treatment\_date columns in ascending order, the first row will give treatment start date of each patient as diagnosis\_date per patients is constant. I will extract the first row and then using difftime() function I will calculate the difference between treatment date and diagnosis date in days.

**Steps:**

We will merge diagnosis\_filtered and Treatment\_filtered datasets by primary key patient\_id.

merge\_diagnosis\_treatment <- merge(diagnosis\_filtered, Treatment\_filtered, by= "patient\_id") #join

Let’s sort the dataframe by patient\_id.

merge\_sorted <- merge\_diagnosis\_treatment[order(merge\_diagnosis\_treatment$patient\_id,merge\_diagnosis\_treatment$treatment\_date),] #sort by patient\_id and treatment\_date

As we are looking for the first-line treatment for each patient, we will keep 1st observation for each patient\_id and remove the details of subsequent treatment visits.

First\_Rows <- merge\_sorted[!duplicated(merge\_sorted$patient\_id), ]

We will calculate the time difference between treatment\_time and diagnosis\_time using difftime() function, we will extract those results in day unit.

First\_Rows$time\_to\_treatment <- difftime(First\_Rows$treatment\_date,First\_Rows$diagnosis\_date, unit="day")  
print(First\_Rows)

## patient\_id diagnosis\_date diagnosis\_code diagnosis treatment\_date  
## 4 2038 2010-01-21 174.9 Breast Cancer 2010-01-24  
## 22 2120 2010-01-09 174.1 Breast Cancer 2010-01-23  
## 53 2175 2010-02-17 174.7 Breast Cancer 2010-02-21  
## 73 2238 2010-01-21 174.9 Breast Cancer 2010-01-21  
## 85 2407 2010-06-13 174.9 Breast Cancer 2010-06-19  
## 117 2425 2010-12-15 174.9 Breast Cancer 2010-12-19  
## 135 2462 2011-01-07 174.9 Breast Cancer 2011-01-11  
## 168 2475 2010-02-17 174.7 Breast Cancer 2010-02-17  
## 174 2607 2010-06-13 174.9 Breast Cancer 2010-07-03  
## 182 2634 2011-02-19 153.9 Colon Cancer 2011-12-20  
## 211 2720 2010-01-09 174.1 Breast Cancer 2010-01-20  
## 218 2735 2011-04-18 174.9 Breast Cancer 2011-04-23  
## 249 2762 2011-01-07 174.9 Breast Cancer 2011-01-10  
## 266 2763 2011-04-19 174.1 Breast Cancer 2011-04-23  
## 281 2770 2011-04-16 153.9 Colon Cancer 2011-04-22  
## 310 3025 2010-12-15 174.9 Breast Cancer 2010-12-21  
## 314 3070 2011-07-25 153.9 Colon Cancer 2011-07-25  
## 327 3095 2011-07-10 153.3 Colon Cancer 2011-07-13  
## 338 3395 2011-10-18 153.3 Colon Cancer 2011-10-18  
## 362 3449 2011-09-09 153.5 Colon Cancer 2011-09-13  
## 366 3749 2011-12-18 153.5 Colon Cancer 2011-12-18  
## 371 3757 2011-10-17 153.4 Colon Cancer 2011-10-22  
## 406 3948 2011-12-18 174.6 Breast Cancer 2011-12-22  
## 448 4057 2012-01-25 153.4 Colon Cancer 2012-01-25  
## 453 4354 2012-02-04 174.8 Breast Cancer 2012-02-09  
## 475 4374 2012-03-20 174.5 Breast Cancer 2012-03-25  
## 489 4692 2012-04-27 174.8 Breast Cancer 2012-04-30  
## 538 5259 2012-05-13 174.3 Breast Cancer 2012-05-17  
## 612 5657 2012-06-07 174.8 Breast Cancer 2012-06-12  
## 632 6281 2012-08-12 174.4 Breast Cancer 2012-08-16  
## 698 6321 2012-09-06 174.2 Breast Cancer 2012-09-10  
## 754 6837 2012-10-20 153.3 Colon Cancer 2012-10-25  
## 791 6840 2012-11-15 153.4 Colon Cancer 2012-11-20  
## 822 6877 2012-12-09 174.3 Breast Cancer 2012-12-03  
## 846 6889 2012-11-17 174.7 Breast Cancer 2012-11-22  
## 864 6922 2012-11-20 174.9 Breast Cancer 2012-11-22  
## 895 7230 2013-01-06 174.9 Breast Cancer 2013-01-09  
## 954 7242 2013-01-17 153.5 Colon Cancer 2013-01-23  
## 991 7796 2013-01-16 174.9 Breast Cancer 2013-01-21  
## 1010 7937 2013-01-06 174.9 Breast Cancer 2013-01-12  
## 1030 7976 2013-03-06 174.1 Breast Cancer 2013-03-11  
## 1039 8480 2013-05-16 174.3 Breast Cancer 2013-05-22  
## 1067 8615 2013-07-18 174.7 Breast Cancer 2013-07-24  
## 1082 8827 2013-07-21 174.9 Breast Cancer 2013-07-18  
## 1091 9331 2013-08-23 174.9 Breast Cancer 2013-08-29  
## 1131 9489 2013-08-19 174.9 Breast Cancer 2013-08-25  
## drug\_code time\_to\_treatment  
## 4 A 3 days  
## 22 B 14 days  
## 53 B 4 days  
## 73 B 0 days  
## 85 A 6 days  
## 117 A 4 days  
## 135 A 4 days  
## 168 B 0 days  
## 174 B 20 days  
## 182 B 304 days  
## 211 B 11 days  
## 218 A 5 days  
## 249 B 3 days  
## 266 A 4 days  
## 281 A 6 days  
## 310 B 6 days  
## 314 D 0 days  
## 327 B 3 days  
## 338 D 0 days  
## 362 B 4 days  
## 366 D 0 days  
## 371 C 5 days  
## 406 A 4 days  
## 448 D 0 days  
## 453 A 5 days  
## 475 C 5 days  
## 489 B 3 days  
## 538 A 4 days  
## 612 B 5 days  
## 632 B 4 days  
## 698 B 4 days  
## 754 B 5 days  
## 791 B 5 days  
## 822 C -6 days  
## 846 B 5 days  
## 864 C 2 days  
## 895 C 3 days  
## 954 B 6 days  
## 991 B 5 days  
## 1010 A 6 days  
## 1030 A 5 days  
## 1039 A 6 days  
## 1067 B 6 days  
## 1082 A -3 days  
## 1091 B 6 days  
## 1131 C 6 days

Let’s check if calculated column time\_to\_treatment is numeric, if false we will convert it to numeric.

is.numeric(First\_Rows$time\_to\_treatment)

## [1] FALSE

First\_Rows$time\_to\_treatment <- as.numeric(as.character(First\_Rows$time\_to\_treatment))

is.numeric(First\_Rows$time\_to\_treatment)

## [1] TRUE

summary\_time<- summary(First\_Rows$time\_to\_treatment)  
print(summary\_time)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -6.0 3.0 4.5 10.8 6.0 304.0

The mean time after being diagnosed to patients start treatment is 10.8 days.

now we will calculate how long it takes for patients to start therapy after being diagnosed for both the cancer types.

colon\_cancer\_df <- First\_Rows[First\_Rows$diagnosis == "Colon Cancer", ]  
breast\_cancer\_df <- First\_Rows[First\_Rows$diagnosis == "Breast Cancer", ]  
colon\_cancer\_df$time\_to\_treatment <- as.numeric(as.character(colon\_cancer\_df$time\_to\_treatment))  
breast\_cancer\_df$time\_to\_treatment <- as.numeric(as.character(breast\_cancer\_df$time\_to\_treatment))

we neglected the extreme values of time\_to\_treatment to make the data consistent.

recheck the format of time\_to\_treatment.

is.numeric(colon\_cancer\_df$time\_to\_treatment)

## [1] TRUE

colon\_cancer\_summary <- summary(colon\_cancer\_df$time\_to\_treatment)  
breast\_cancer\_summary <- summary(breast\_cancer\_df$time\_to\_treatment)  
print(colon\_cancer\_summary)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00 0.00 4.50 28.17 5.25 304.00

print(breast\_cancer\_summary)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -6.000 3.250 4.500 4.676 6.000 20.000

The colon cancer patients mean time to get the treatment after diagnosis is 28.17 days and that of breast\_cancer patients is 4.7 days.

### Question 3

1. A patient’s first-line treatment is the drug (i.e., monotherapy) or set of drugs (i.e., combination therapy) that the patient received at the start of systemic treatment (e.g.,first treatment after earliest diagnosis) for their disease (for more information on first-line treatments, click here). Without access to information about the clinician’s specific decision making, we can infer a patient’s first-line treatment regimen based on the drug or set of drugs they received in their first treatment instance. Using this approach, which treatment regimens [i.e., drug(s)] do you think would be indicateds as first-line treatment for patients with… ○ breast cancer only? ○ colon cancer only? ○ both breast and colon cancer?

Ans: To know the first-line treatment regimen, I thought of counting the occurrences of drug code within each cancer type during first treatment\_date after diagnosis.

**Steps:**

Count the occurrences of each drug within each cancer type

drug\_counts <- First\_Rows %>%  
 group\_by(First\_Rows$diagnosis, First\_Rows$drug\_code) %>%  
 summarise(count = n()) %>%  
 ungroup()

## `summarise()` has grouped output by 'First\_Rows$diagnosis'. You can override  
## using the `.groups` argument.

print(drug\_counts)

## # A tibble: 7 × 3  
## `First\_Rows$diagnosis` `First\_Rows$drug\_code` count  
## <chr> <chr> <int>  
## 1 Breast Cancer A 13  
## 2 Breast Cancer B 16  
## 3 Breast Cancer C 5  
## 4 Colon Cancer A 1  
## 5 Colon Cancer B 6  
## 6 Colon Cancer C 1  
## 7 Colon Cancer D 4

From above table we could observe that A and B are preferred first-line treatment regimen for Breast Cancer, while B and D are preferred first-line treatment regimen for Colon Cancer. Drug C is used in Breast cancer treatment as well.

The breast cancer first-line treatment regimen include surgery followed by radiation and chemotherapy. Chemotherapy for breast cancer can help slow the growth of cancer, shrink a tumor, and increase the chance of curing the condition. Adjuvant therapy refers to chemotherapy treatment after surgery to remove any remaining cancer cells and reduce the likelihood of the cancer regrowing. Neoadjuvant refers to treatment with chemotherapy before surgery to shrink the tumor and increase the odds of surgical success.When cancer metastasizes, or spreads, outside of the breast tissue, doctors may recommend chemotherapy as the main treatment. axanes, such as paclitaxel and docetaxel, cyclophosphamide carboplatin, anthracyclines, such as epirubicin and doxorubicin, 5-fluorouracil or capecitabine these are the drugs used in the first-line treatment of breast cancer.for more information click here [Link](https://www.medicalnewstoday.com/articles/chemotherapy-for-breast-cancer#chemotherapy-regimens-for-breast-cancer)

People with colon cancers that have not spread to distant sites usually have surgery as the main or first treatment. Chemotherapy may also be used after surgery (called adjuvant treatment). Most adjuvant treatment is given for about 3 to 6 months.For colon cancer, typical first line therapies include Capecitabine, Fluorouracil, Irinotecan, and combinations of these and other drugs. Pembrolizumab is also a first-line treatment for some unresectable colon cancers.For more information please visit[Link](https://www.cancer.org/cancer/types/colon-rectal-cancer/treating/by-stage-colon.html#:~:text=People%20with%20colon%20cancers%20that,about%203%20to%206%20months.)