Abstract:

In this paper, we examine how hepatocellular carcinoma patients react to treatment under sorafenib. The dataset analyzed in this study contains results from 2016-2019; the total number of participants in this study is 18,177: 9,689 female and 8,488 male. The dataset is specific to Canada and includes participants from Alberta (AB), the Atlantic region, British Columbia (BC), Ontario (ON), the Prairies region, Quebec (QC), and undeclared regions (UNKWN); while age groups are segmented in this study for organizational purposes, the dataset examines participants that are aged 18 years or older.

    In this paper, we examine how sex, region, age, and whether participants are using another cancer treatment - which will be called 'con act' throughout this study - affects participation in treatment and success rates in using sorafenib. Due to the study's multivariable nature, we decided to use con act as the independent variable and examine age, sex, and region against it. While it may seem intuitive to use age as the independent variable, the age results were very biased and volatile depending on the age group examined; the same behaviour was present in the sex variable. Con act was the variable with the least fluctuation and bias when measured with respect to all the other variables; as such, our team decided to use it as the basis for our analysis.

    While the data segmented participants based on three measures - Tx, events, and censored - our analysis revealed that the events and censored measures were too infrequent and random to draw any conclusions from. A great majority of the participants - roughly 92% - were strictly on the Tx measure throughout the study with only 8% switching to the events and censored measures during the course of the treatment. As such, our analysis groups all the measures into one large sum and treats them as one. Although this might seem counterintuitive, our analysis shows no variations or patterns between the measures with respect to the other variables.

    As mentioned above, this paper explores factors that may influence a patient's successful therapy and the potential reasons for discontinuing therapy with sorafenib. Factors that might predict successful therapy are lower age, solely using sorafenib, or not being a female that is older than forty five years. Potential reasons for discontinuing therapy were determined to be

death, the patient deciding to stop treatment, fully recovering from the disease, or seeking alternative treatment non-concurrently with sorafenib.

According to our analysis, the total number of patients that remain on treatment for at least 9 months was 8,784, or 48.3% of the total number of participants; furthermore, the monthly rate of discontinuation for all patients turned out to be roughly 12% with about 36 patients discontinuing treatment every month. The relationships discussed in the paper can be applied on a province-to-province basis since most, if not all provinces, share the exact same patterns. Future papers may look to consider other factors that lead to discontinuation of treatment such as underlying health conditions, racial background, and the number of offspring the participant has.

Methods:

Due to the large and detailed nature of the data set used, our analysis required extensive filtering and visualization techniques. To meet these requirements, a python program was developed to index through the data set, query the needed information, and visualize the trends and patterns observed. To read and extract the information needed, the openpyxl library was used; however, even though the data was extracted, it was represented crudely within the python environment. As such, the pandas framework was used to organize the data into a format that is more pleasant to interact with and view.

Pandas was not only used for data representation but also for querying and filtering the information. The data was queried using boolean logic and python's internal mathematical operators. Although the use of openpyxl and pandas was extremely helpful in representing, organizing, and querying the data, it proved difficult to use mathematical operators on a data set that extends for 40 months. As such, we used the numpy library to take advantage of linear algebra arithmetic to simplify our calculations. When extracting the data for a certain set of variables, we used numpy to represent that demographic's information as a vector - a vector that has 40 elements; as such, it was extremely simple to calculate the averages, summations, rates, and correlations between variables.

Lastly, the matplotlib library was used to visualize the data and represent it into an intuitive format. While the data was mainly represented in bar graphs, matplotlib's features allowed us to segment the information, view it with a corresponding data table, and highlight important patterns and features that belong to certain variables.

Discussion:

Firstly, we will evaluate how many patients stay on treatment for at least 9 months, and explore possible factors leading to successful therapy. [Figure 1](#) tells us that there are 4,567 males and 4,217 females who stay on treatment for at least 9 months. This brings the total number of patients in all age groups on treatment for at least 9 months, nationwide, to 8,784. Next, [Figures 2-4](#) show the total number of patients on therapy for at least 9 months for age groups: 18-19, 35-39, and 65+ respectively. There are 18 patients between the ages of 18-19 on therapy for that period of time ([Figure 2](#)). This compared to the 512 patients aged 35-39 ([Figure 3](#)) and 730 patients aged 65+ ([Figure 4](#)) shows an upward trend with patients currently on treatment. As age increases, so does the likelihood of being on treatment for a longer time. Being on treatment longer is a good indication that the disease is still persistent in the patient's body. The longer the disease is in a patient's body, the more likely health complications may occur. As per [figures 2-4](#), a lower age tends to result in less therapy time which is likely due to successful treatment. Hence, the lower a person's age, the more likely they are to have successful therapy.

Secondly, comparing [Figure 5](#) with [Figure 6](#) shows that being on another cancer treatment, concurrently with sorafenib, doesn't correspond to higher probabilities of successful therapy. In fact, it seems as though taking sorafenib leads to a higher chance of successful therapy. [Figure 5](#) shows us that in the first 8 months of treatment 6,919 out of 9,393 patients choose not to pursue additional therapy concurrently with sorafenib. [Figure 6](#) shows us that in the last 8 months of treatment 246 out of 332 patients are also not doing any additional therapy. So, roughly 73.66% of patients in the first 8 months of treatment do not do additional therapy, and 74.1% of patients do not do additional therapy in their last 8 months of treatment. The minimal change here indicates that as treatment time increases, there is *not* a motive on the patient's side to seek additional therapy. The main reason for this is likely because the patient is seeing great results from their treatment on sorafenib. By the time a patient is in their last 8 months of treatment, they have spent almost 30 months taking some form of anti-cancer therapy. Meaning, if they were not seeing signs of progress, the patient would likely seek alternative therapies. Additionally, it seems that any side-effects produced by taking sorafenib did not seem to cause detrimental damage to the patient's health. Therefore, alternative therapies taken concurrently with sorafenib do not seem to influence the likelihood of successful therapy. With

that, we can conclude that sorafenib, whether taken individually or concurrently with another treatment, will likely result in successful therapy.

Thirdly, it seems as though gender plays a role on whether therapy will be successful or not, however, this is contingent on one's age. Figure 4 and Figure 7 show that there is a tendency for older females to be on treatment longer than males the same age. Comparing Figure 3 to Figure 4 and Figure 7 respectively shows a jump in the number of females who tend to be on treatment for a longer time. This could be due to two main reasons. First, it's possible that the disease itself takes a higher health toll on older females (aged 45+). Second, it's possible that older females do not respond as strongly to sorafenib when compared to other demographics of patients. As discussed before, the longer the cancer is present in a patient's body, the more likely health complications may occur. Hence, since older females tend to stay on treatment longer than other demographics, we may conclude that *not* being an older female will likely result in successful treatment.

Next, we will evaluate the monthly rate of discontinuation, and assess possible factors leading to the discontinuation of therapy. Figure 8 shows us the monthly discontinuation rate for all patients. Averaging the monthly discontinuation rate for each demographic leaves us with an overall monthly discontinuation rate of 12%. Meaning, roughly 12% of patients discontinue treatment every month. Firstly, patients may discontinue treatment due to death or the cancer reaching a terminal phase. There are a couple of factors which may lead to the death of a patient. First, death could be due to the cancer reaching a terminal phase that ends the patient's life. Second, death may have arisen due to the side effects of sorafenib or alternative therapies. These side effects may have caused the patient to develop further health complications that end the patient's life. Third, the death could be due to natural reasons independent of the patient's cancer. It's very possible that a heart attack, not induced by the patient's cancer, was the reason for their passing. Figure 10 shows a large increase of discontinuation of treatment between quarter 3 and quarter 4, specifically for males aged 65+. 46 out of 52 did not make it from quarter 3 to quarter 4. That's a staggering 88.5% quarterly discontinuation rate. By the time a patient is at quarter 3, they have been on treatment for roughly 20 months. And by the time a patient is at quarter 4, they have been on treatment for roughly 30 months. This prolonged treatment is not necessarily a good sign a patient is getting better. Moreover, since Figure 10 focuses on the 65+ age group, it is highly likely that the majority of the 88.5% of males who

discontinued treatment between quarter 3 and quarter 4 did so due to their death. Moreover, some subset of the 88.5% of males 65+ who discontinued treatment likely did so due to the cancer reaching a terminal phase. The terminal phase of any cancer disease, and in our case liver cancer, means that there is no chance of the patient getting any better given any treatment. Hence, the patient is left with the choice of continuing treatment and hoping for a miracle, or stopping treatment completely. Some majority of the 88.5% of males may have decided to discontinue treatment since they reached this terminal stage. Therefore, both death and reaching the terminal stage of hepatocellular carcinoma are possible reasons for discontinuation of therapy.

Secondly, patients who fully recover from the disease will most certainly discontinue treatment. Figure 9 shows how patients ages 18-19 do not even make it past the quarter 3 mark for their treatment. As discussed previously, a lower age and less time on treatment tends to result in successful therapy. Figure 9 indicates both of these. 36 out of the original 40 female patients aged 18-19 did not even make it to quarter 2 of their treatment. So, 90% of females aged 18-19 only required 9 months or less for therapy. This most certainly indicates that the majority of those females had a full-recovery. There may of course be other factors such as death which played a role in the 90% quarterly discontinuity, however, the more likely reason is a full-recovery due to the age (low), length of time on therapy (low), and demography (young female) of the patient.

Thirdly, we will consider patients who choose to take an alternative treatment non-concurrently with sorafenib. Figure 8 shows a large increase in the discontinuation of therapy between quarter 1 and quarter 2. Roughly 44% of all patients who began taking sorafenib ended up discontinuing sometime in the first 9 months of their treatment. A likely reason behind this is that the patient found an alternative therapy to sorafenib. This alternative therapy could have required the patient to stop taking sorafenib all together due to side effects of taking the two at the same time. Or, the patient simply decided to switch to a new therapy all together due it being more effective than sorafenib. Since there is no clear cure to cancer, patients diagnosed with it tend to experiment with various treatments. Hence, these reasons may apply to a subset of the 44% of patients who discontinued treatment between quarter 1 and quarter 2. If so, we can conclude that patients who choose to take an alternative treatment non-concurrently with sorafenib may be a potential reason for discontinuing therapy (with sorafenib).

Appendix:

Here is a link to our github repository. It includes our python program and the dataset provided.

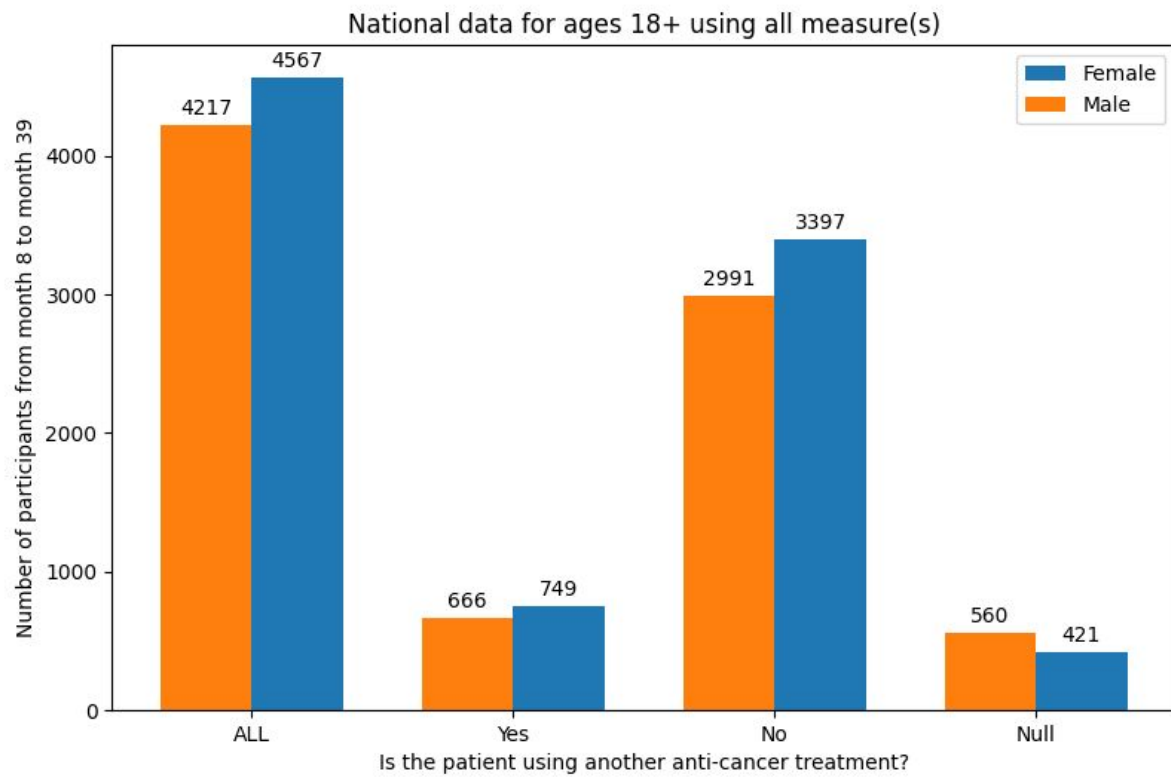Figure 1: Patients on treatment for over 9 months

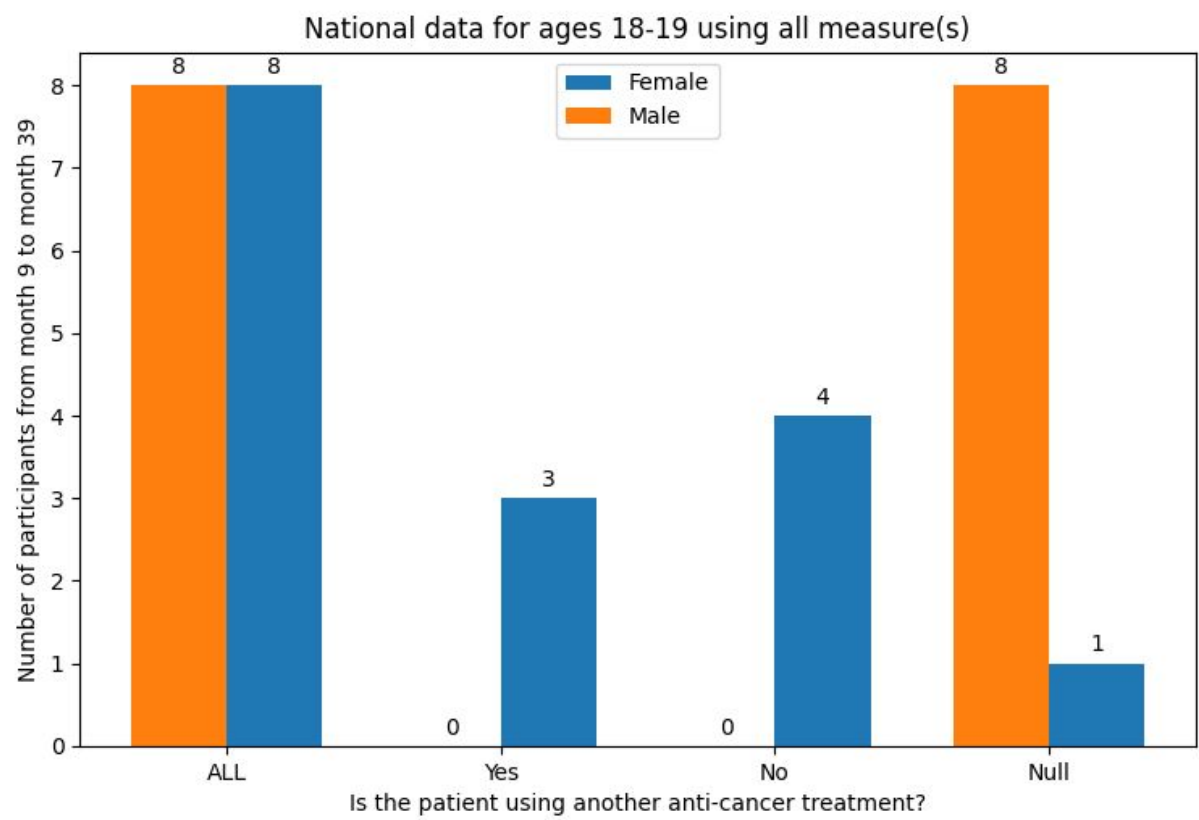Figure 2: Patients aged 18-19 on treatment for over 9 months

National data for ages 18-19 using all measure(s)

Figure 3: Patients aged 35-39 on treatment for over 9 months
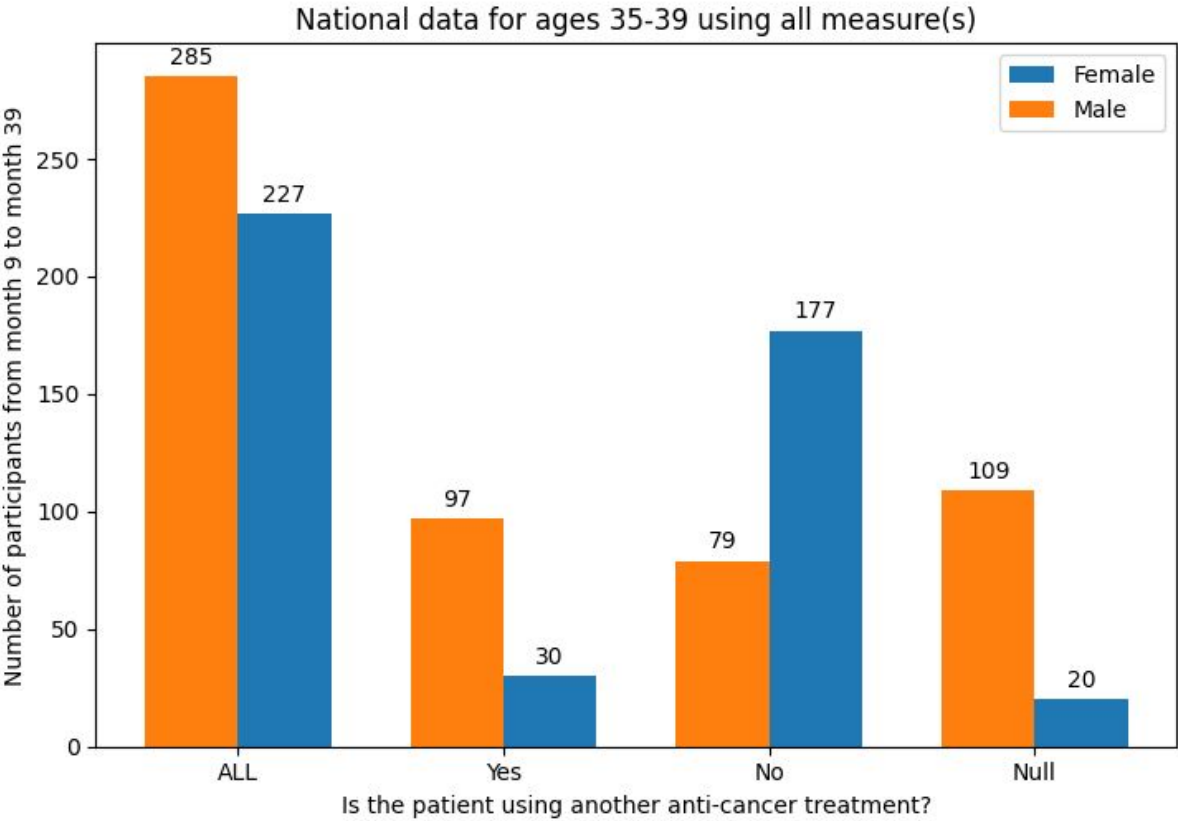

National data for ages 35-39 using all measure(s)

Figure 4: Patients aged 65+ on treatment for over 9 months
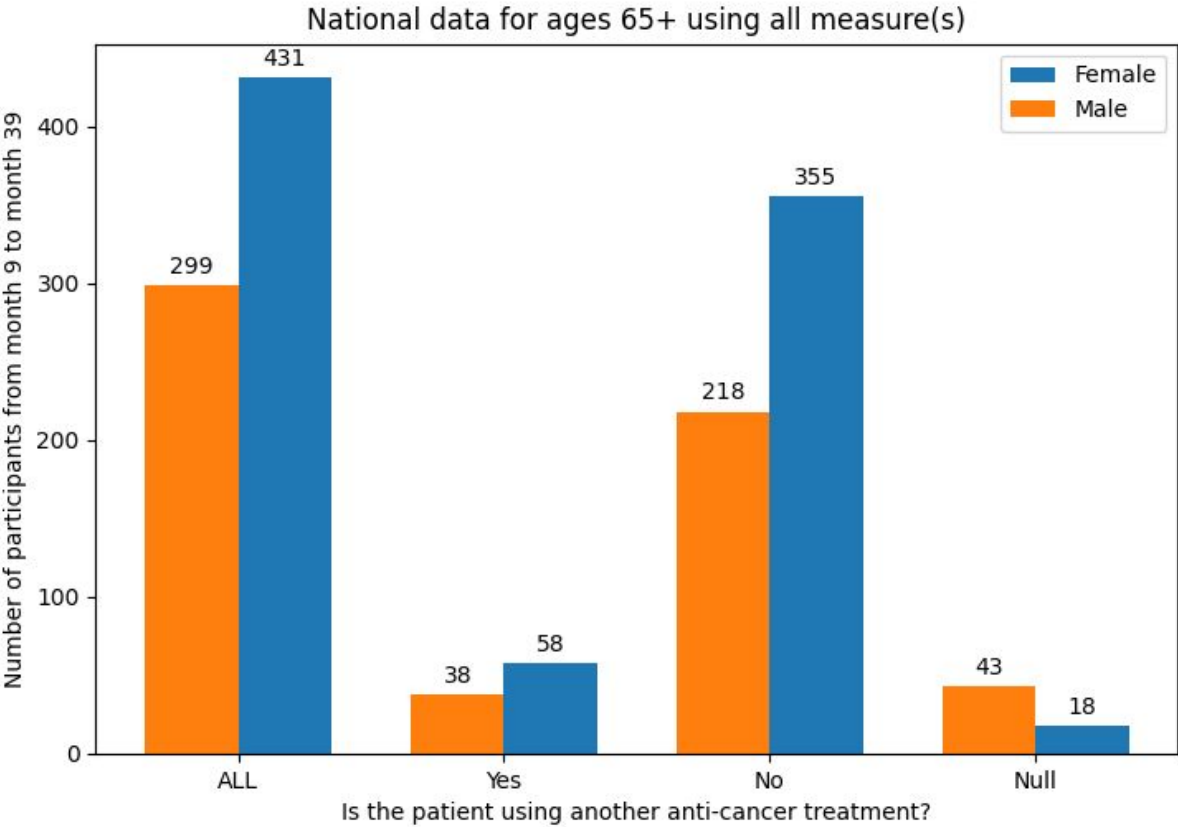
Figure 5: Patients on treatment for first 8 months

National data for ages 18+ using all measure(s)
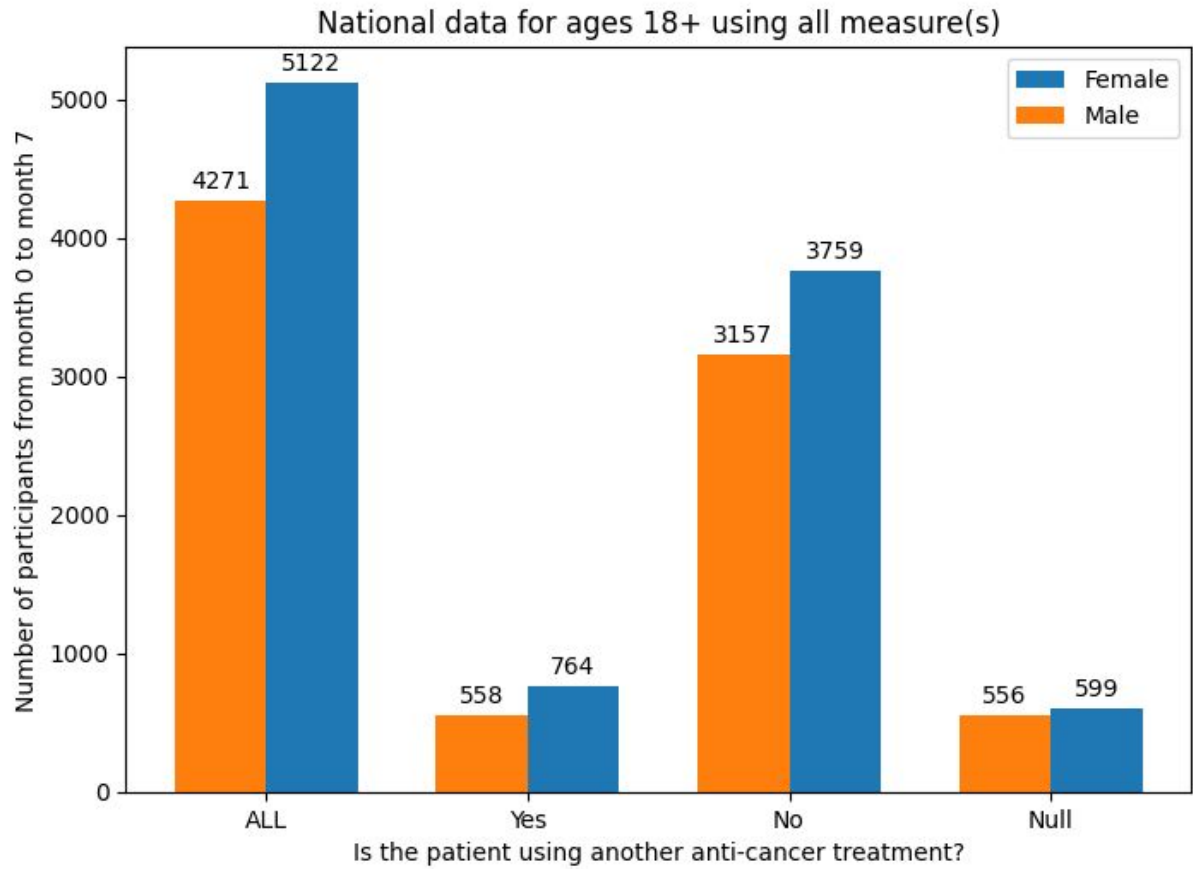
Figure 6: Patients on treatment for last 8 months

National data for ages 18+ using all measure(s)
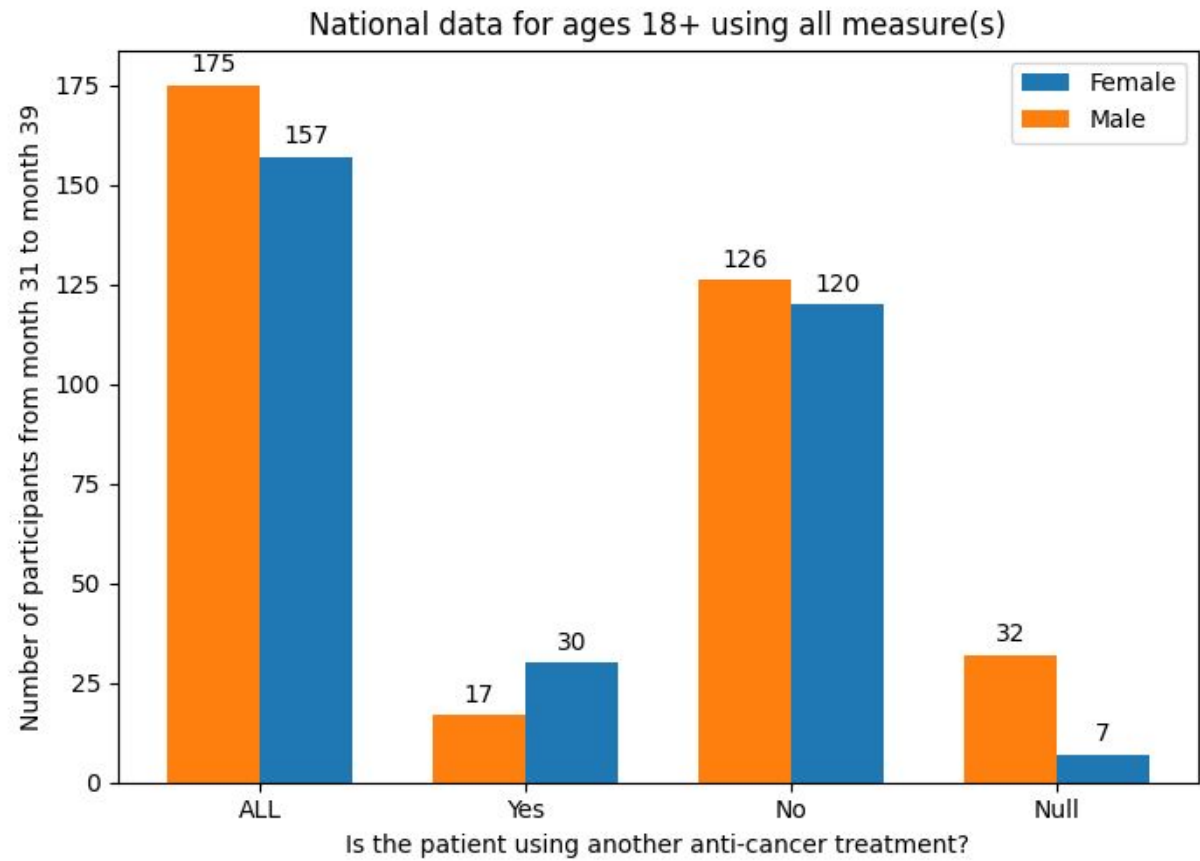
Figure 7: Patients aged 45-49 on treatment for at least 9 months
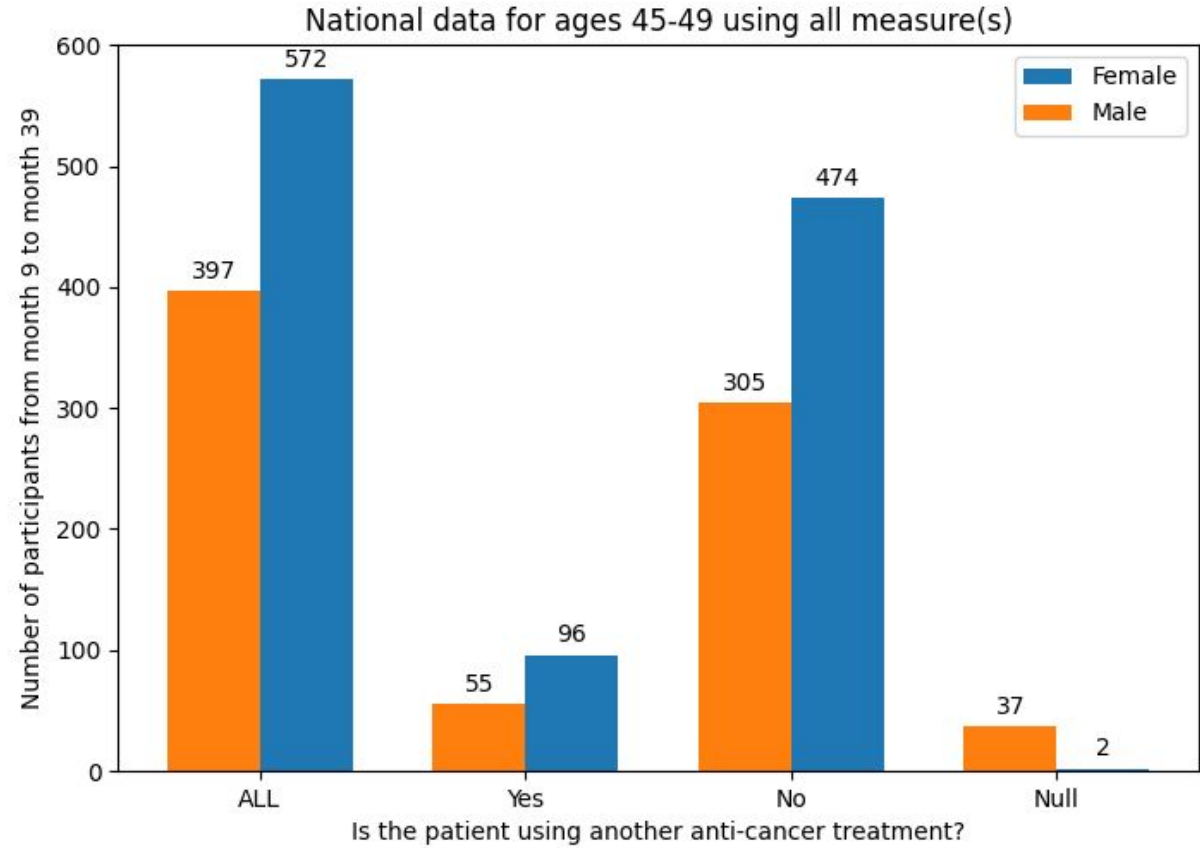
Figure 8: Monthly Discontinuation Rate for all ages

National Quarter Data and Monthly Discontinuation rate for ages 18+

	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Monthly Rate
Female No	4399.0	1866.0	745.0	146.0	12.1
Female Yes	893.0	412.0	170.0	38.0	12.8
Female Null	686.0	268.0	58.0	8.0	12.7
Male No	3698.0	1603.0	689.0	158.0	12.6
Male Yes	663.0	389.0	149.0	23.0	12.3
Male Null	653.0	307.0	118.0	38.0	9.1
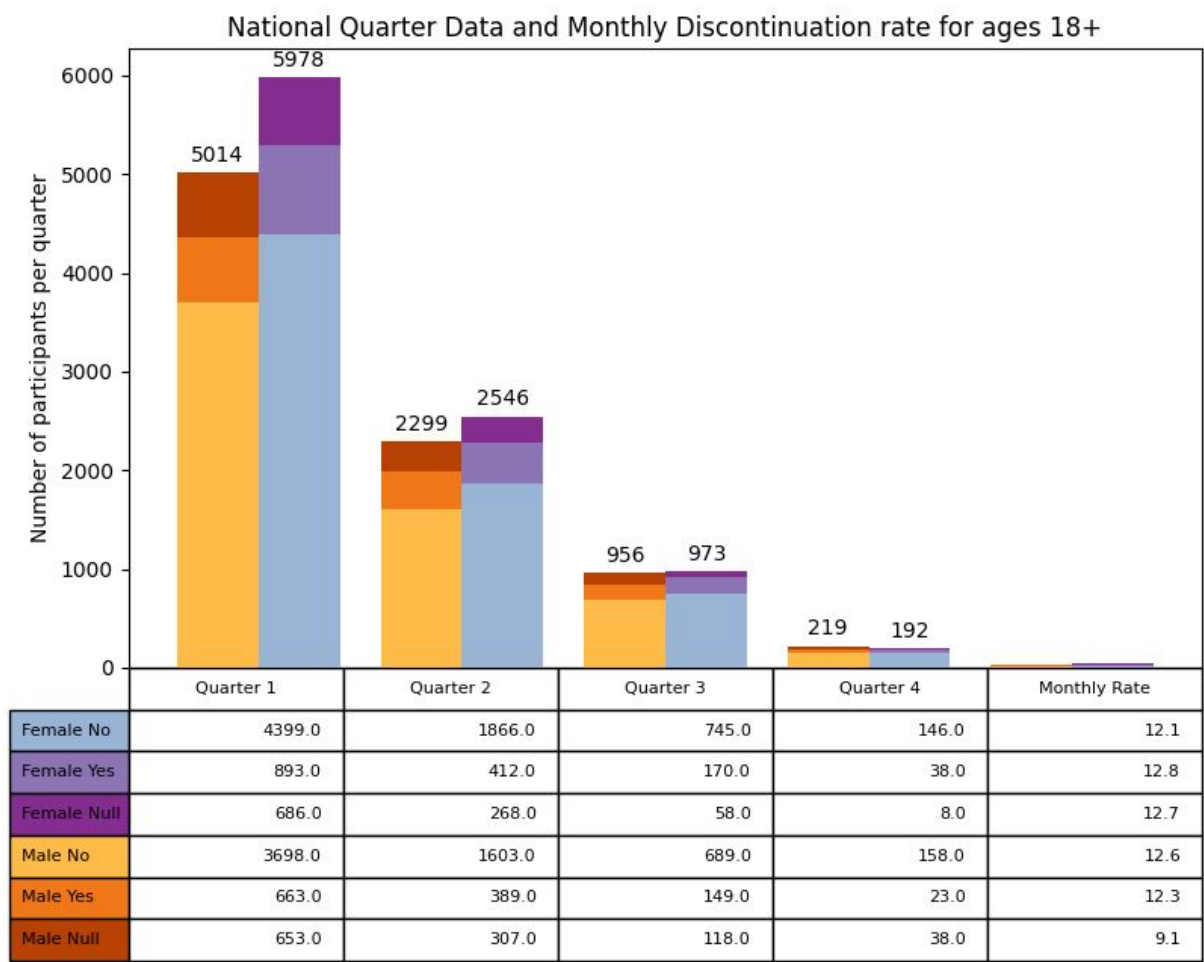
Figure 9: Monthly Discontinuation Rate for patients aged 18-19

National Quarter Data and Monthly Discontinuation rate for ages 18-19

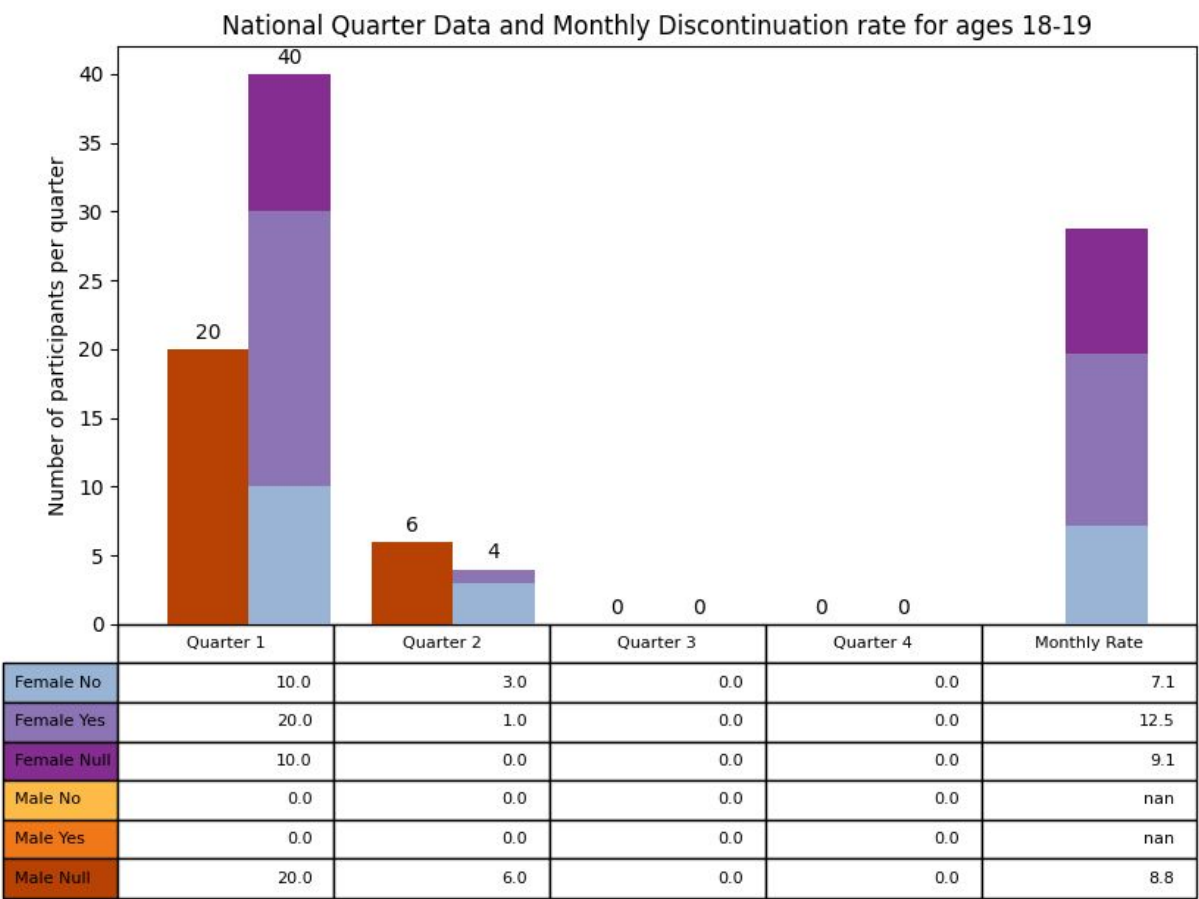|	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Monthly Rate
Female No	10.0	3.0	0.0	0.0	7.1
Female Yes	20.0	1.0	0.0	0.0	12.5
Female Null	10.0	0.0	0.0	0.0	9.1
Male No	0.0	0.0	0.0	0.0	nan
Male Yes	0.0	0.0	0.0	0.0	nan
Male Null	20.0	6.0	0.0	0.0	8.8

Figure 10: Monthly Discontinuation Rate for patients aged 65+

National Quarter Data and Monthly Discontinuation rate for ages 65+

	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Monthly Rate
Female No	529.0	217.0	75.0	23.0	9.0
Female Yes	64.0	39.0	14.0	0.0	9.1
Female Null	37.0	10.0	7.0	0.0	7.6
Male No	434.0	140.0	45.0	6.0	12.0
Male Yes	86.0	32.0	0.0	0.0	11.6
Male Null	73.0	30.0	7.0	0.0	9.9