Research Question……………………………………………………………………….2

Introduction………………………………………………………………………………..2

Aims of the Exploration…………………………………………………………………..3

Rationale…………………………………………………………………………………..3

PART A- BAYES' THEOREM……………………………………………………………...4

1. Test 1.1..…………………………………………………………………………...4
2. Test 1.2……………………………………………………………………………6
3. Test 1.3……………………………………………………………………………8
4. Test 1.4…………………………………………………………………………...9
5. Test 1.5..…………………………………………………………………………10

Results for PART A………………………………………………………………………..12

PART B- MARKOV CHAINS……………………………………………………………...12

1. Test 2.1..…………………………………………………………………………13
2. Test 2.2…………………………………………………………………………...14
3. Test 2.3…………………………………………………………………………...15
4. Test 2.4…..………………………………………………………………………16
5. Test 2.5..…………………………………………………………………………17

Results for PART B………………………………………………………………………..18

Conclusion…………………………………………………………………………………18

Limitations of the Exploration……………………………………………………………..19

References…………………………………………………………………………………19

Research Question:

What is the probability of the accuracy of the diagnosis made by certain medical tests for detecting different types of cancers, and the probability of cancer occurrence in the future?

Introduction:

Throughout the entirety of my school career, Biology has always been my favourite subject. My fascination with learning about the human physiology defined some of my after-school activities, such as watching medical dramas as a hobby.

In one particular episode of a medical show, a patient at the hospital had been wrongly diagnosed with cancer, leading to the patient suing the hospital. I realized that the consequences of a misdiagnosis can be severe. The treatment that the patient receives can be inappropriate, and they may also be left untreated. Moreover, the mental health of the patient may be affected.

This led me to wonder the extent to which a diagnostic test can be wrong and the extent to which they can be relied upon. I then conducted some research using reliable sources like NCBI or PubMed, to find the rate at which certain cancer tests give the wrong diagnosis in terms of false negative rates, i.e., the frequency with which cancer is not detected even though the patient has cancer, and false positive rates, i.e., the frequency with which a patient has been diagnosed with cancer even though they are not affected. I was also led to question whether this data could predict cancer occurrence in the following years of the studies from which the data was collected.

Aims of the Exploration:

There are two aims of this exploration. The first (PART A) is to investigate the accuracy of a positive result given by certain medical tests that are used to detect breast, lung, prostrate, pancreatic and bladder cancer, i.e., the rate of a patient having the disease given a positive result. The second (PART B) is to calculate the probability of cancer occurrence following two years of the study from which false positive and false negative rates were taken, given that false positive and false negative rates stay the same and the medical test in question is used for the diagnosis.

Rationale:

The raw data collected to be used to perform calculations will be the false positive rate and false negative rate of the test, along will the probability that the patient has the disease that the specific test is designed to detect. The first aim will be assessed by using the Bayes theorem:

$$P(\mathrm{A}/\mathrm{B}) = \frac{P(\mathrm{B}/\mathrm{A}) \times P(\mathrm{A})}{P(\mathrm{B})}$$

B will stand for the probability of a positive test, while A will stand for the probability that the patient has contracted the disease. Therefore, P(A/B) is the probability that the patient has the disease when given a positive result. For B, the false positive and true positive rates need to be added.

A Markov chain is a system that undergoes changes from one state to another with reference to certain probabilistic rules. The second aim will be assessed using Markov Chains. I decided to use Markov Chains because I found that it was the most reliable way to make future predictions for cancer occurrences. In this investigation, the current state refers to the probability that the patient will test positive in a particular year, and the next state refers to the probability that the patient will test negative the same year.

All calculations are shown correct to three significant figures, and all calculations were done using a Casio Graphic Display Calculator.

PART A- BAYES' THEOREM

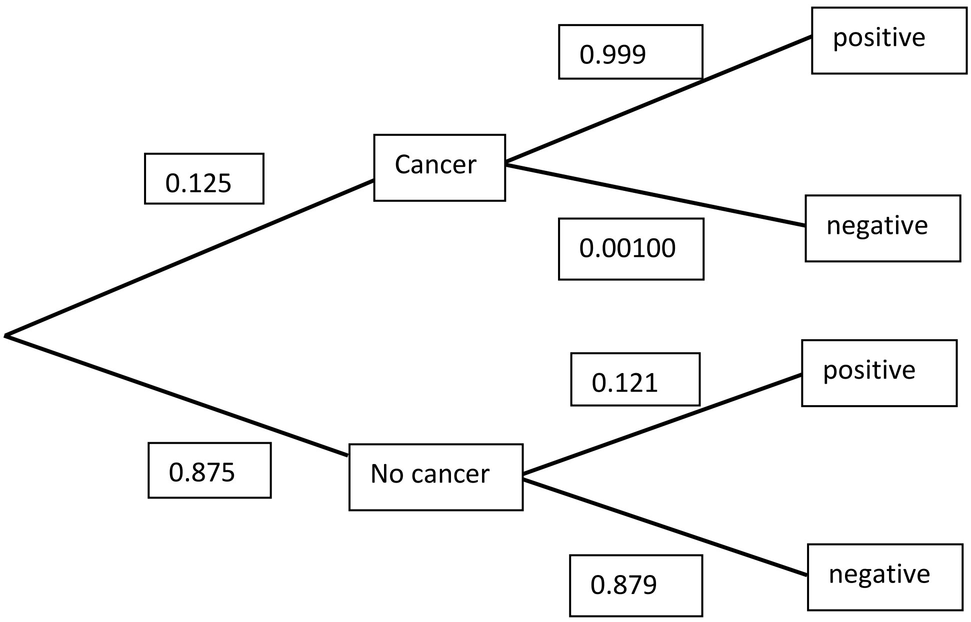
Test 1.1- Accuracy of diagnosis of breast cancer from digital mammography screening

The probability that a woman belonging to the United States has breast cancer is **12.5%**[1]. A study[2] showed that the false positive rate of digital mammography screening was 121.1/1000, i.e., **12.11%** and the false negative rate of digital mammography screening was 1.0 to 1.5/1000. I calculated the mean of 1.0 and 1.5, 1.25, and divided it by 1000 giving me a false negative rate of **0.125%**.

[1] "U.S. Breast Cancer Statistics." *Breastcancer.org*, 4 Feb. 2021, www.breastcancer.org/symptoms/understand_bc/statistics.

[2] Nelson, Heidi D, et al. "Factors Associated With Rates of False-Positive and False-Negative Results From Digital Mammography Screening: An Analysis of Registry Data." *Annals of Internal Medicine*, U.S. National Library of Medicine, 16 Feb. 2016, www.ncbi.nlm.nih.gov/pmc/articles/PMC5091936/.

Figure 1: Tree diagram for probability of accuracy of breast cancer diagnosis



$$P(\text{Cancer/positive result}) = \frac{P(\text{positive result/Cancer}) \times P(\text{Cancer})}{P(\text{positive result})}$$

$$\frac{0.999 \times 0.125}{(0.125 \times 0.999) + (0.875 \times 0.121)} = 0.541$$

Test 1.2- Accuracy of diagnosis of lung cancer using CT-Guided Automated Needle Biopsy of Lung Nodules

A study[3] carried out a biopsy (cancer test) on 123 patients. 3/123 had a false positive result, leading to a false positive rate of **2.44%**. On the other hand, 21/123 patients had a false negative result, causing a false negative rate of **17.1%**. The probability that a man contract lung cancer in his lifetime is nearly **1 in 15** and **1 in 17**[4] for a woman.

$$\frac{1}{15} + \frac{1}{17} = \frac{32}{255} = 0.125$$

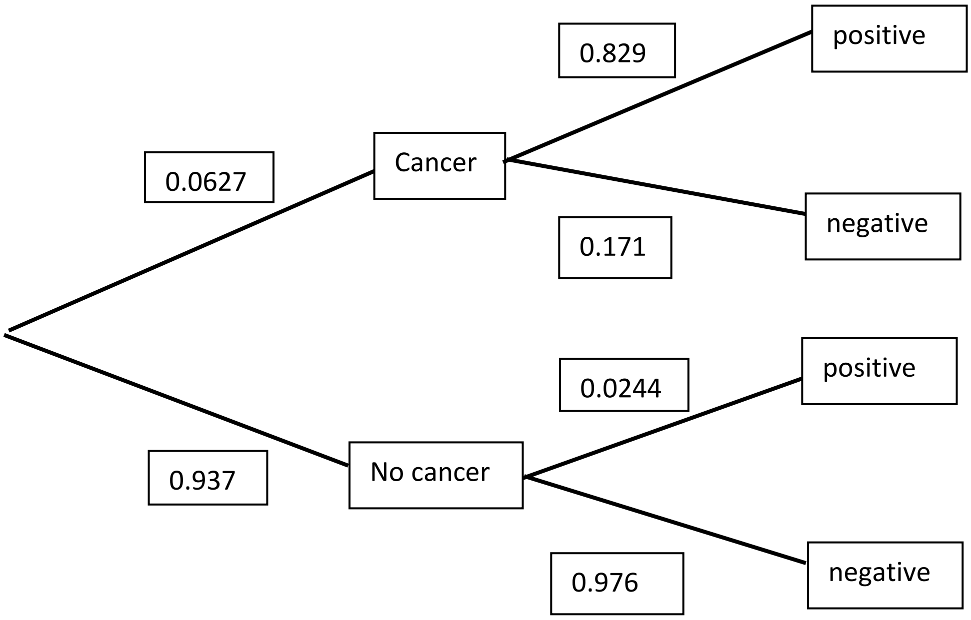
$$\frac{0.125}{2} = 0.0627$$

$0.0627 \times 100 = 6.27\%$ probability that the patient has lung cancer.

[3] Tsudaka, Hiroshi, et al. Diagnostic Accuracy of CT-Guided Automated Needle Biopsy of Lung Nodules. July 2000, www.ajronline.org/doi/pdf/10.2214/ajr.175.1.1750239.

[4] “Lung Cancer Statistics: How Common Is Lung Cancer.” *American Cancer Society*, www.cancer.org/cancer/lung-cancer/about/key-statistics.html.

Figure 2: Tree diagram for probability of accuracy of lung cancer diagnosis



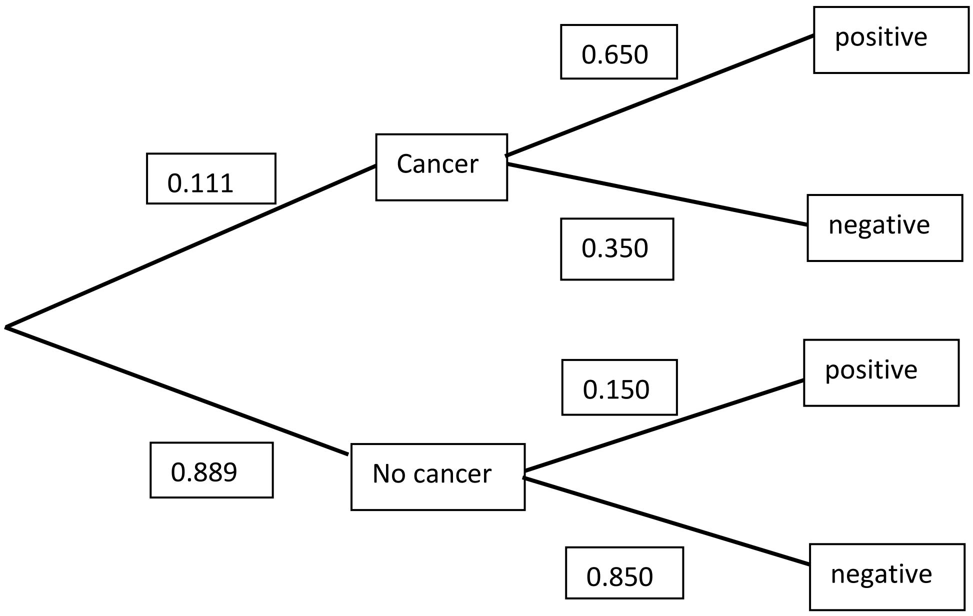
$$P(\text{Cancer/positive result}) = \frac{P(\text{positive result/Cancer}) \times P(\text{Cancer})}{P(\text{positive result})}$$

$$\frac{0.829 \times 0.0627}{(\ 0.829 \times 0.0627\ ) + (\ 0.937 \times 0.0244\ )} = 0.695$$

Test 1.3- Accuracy of diagnosis of prostate cancer using prostate-specific antigen (PSA) testing and transrectal ultrasound-guided (TRUS) prostate biopsy

PSA testing has a **15%** false positive rate, and a **35%** false negative rate[5]. Nearly 1 in 9 men contract prostate cancer during their lifetime[6]. This gives us a probability of **11.1%** that the patient will have the cancer.

Figure 3: Tree diagram for probability of accuracy of prostate cancer diagnosis



$$P(\text{Cancer/positive result}) = \frac{P(\text{positive result/Cancer}) \times P(\text{Cancer})}{P(\text{positive result})}$$

$$\frac{0.650 \times 0.111}{(\ 0.650 \times 0.111\ ) + (\ 0.889 \times 0.150\ )} = 0.351$$

[5] Krasnow, Ross E., and Lambros Stamatakis. “What to Do When Prostate Cancer Biopsy/PSA Test Results Conflict.” *MedStar Washington Hospital Center Blog Center View*, 28 Sept. 2019, blog.medstarwashington.org/2017/09/28/prostate-cancer-high-psa-negative-biopsy/.

[6] “Key Statistics for Prostate Cancer: Prostate Cancer Facts.” *American Cancer Society*, www.cancer.org/cancer/prostate-cancer/about/key-statistics.html.

Test 1.4- Accuracy of the diagnosis of pancreatic cancer using ultrasonography

A study[7] was carried out on 184 patients that were suspected of having pancreatic cancer. When the results came back, the false negative rate was **33%** and the false positive rate was **28%**. On a worldwide scale, the incidence of pancreatic cancer is **5.5 per 100,000** for men, on the other hand it is **4.0 per 100,000** for women[8].

$$\frac{4}{100000} + \frac{5.5}{100000} = \frac{9.5}{100000} = 0.000095$$

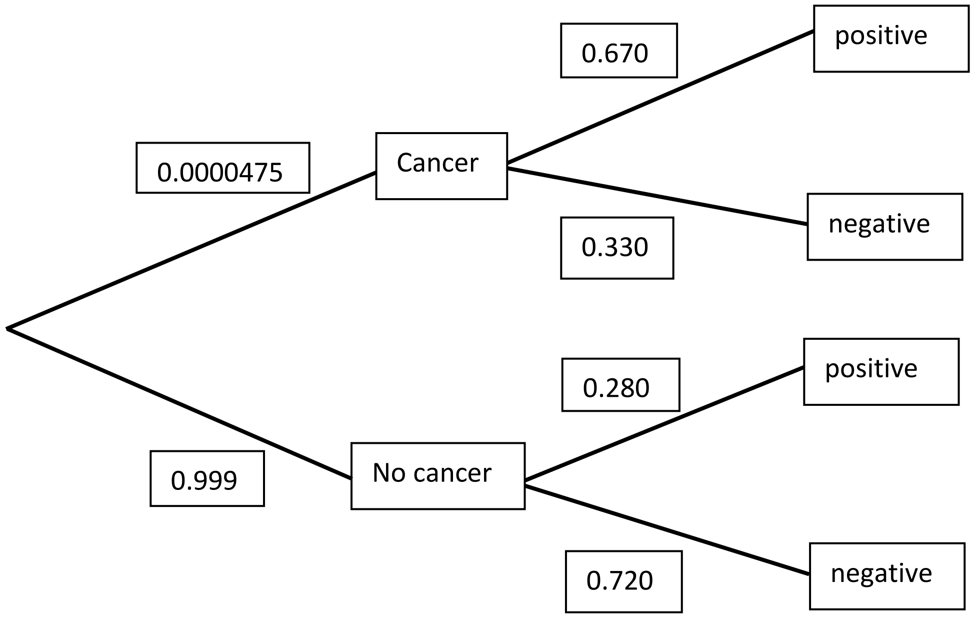
$$\frac{0.000095}{2} = 0.0000475$$

$0.000045 \times 100 = 0.00475\%$ probability of patient having pancreatic cancer.

[7] Fitzgerald, P J, et al. “The Value of Diagnostic Aids in Detecting Pancreas Cancer.” *Cancer*, U.S. National Library of Medicine, Mar. 1978, pubmed.ncbi.nlm.nih.gov/638974/.

[8] Rawla, Prashanth, et al. “Epidemiology of Pancreatic Cancer: Global Trends, Etiology and Risk Factors.” *World Journal of Oncology*, Elmer Press, Feb. 2019, www.ncbi.nlm.nih.gov/pmc/articles/PMC6396775/.

<u>Figure 4: Tree diagram for probability of accuracy of pancreatic cancer diagnosis</u>



$$P(\text{Cancer/positive result}) = \frac{P(\text{positive result/Cancer}) \times P(\text{Cancer})}{P(\text{positive result})}$$

$$\frac{0.670 \times 0.0000475}{(\ 0.670 \times 0.0000475) + (\ 0.999 \times 0.280\ )} = 0.000114$$

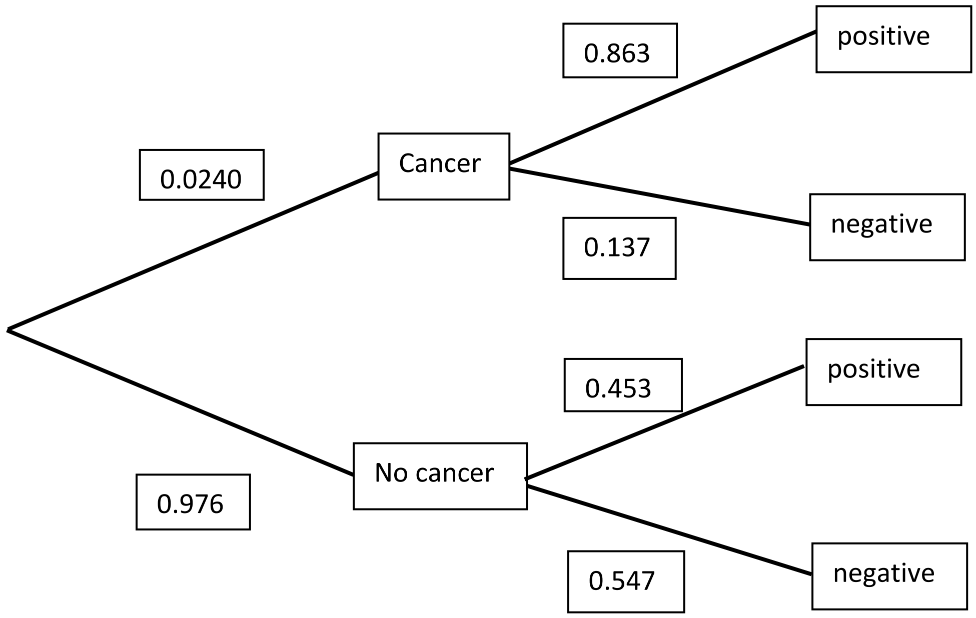
Test 1.5- Accuracy of the diagnosis of bladder cancer using CT urography

A study[9] was carried out on 1623 patients that were suspected to have bladder cancer. Cancer was detected in 95 of these patients, among which there were 43 false positives and 13 false negatives. According to this information, the false

[9]Trinh ,Tony W, et al. “Bladder Cancer Diagnosis with CT Urography: Test Characteristics and Reasons for False-Positive and False-Negative Results.” *Abdominal Radiology (New York)*, U.S. National Library of Medicine, Mar. 2018, pubmed.ncbi.nlm.nih.gov/28677000/.

negative rate is **13.7%** and the false positive rate is **45.3%**. About **2.4%** of men and women will contract bladder cancer in their lifetime.[10]

Figure 5: Tree diagram for probability of accuracy of bladder cancer diagnosis



$$P(\text{Cancer}/\text{positive result}) = \frac{P(\text{positive result}/\text{Cancer}) \times P(\text{Cancer})}{P(\text{positive result})}$$

$$\frac{0.863 \times 0.0240}{(\ 0.863 \times 0.0240) + (\ 0.976 \times 0.453\ )} = 0.0447$$

[10]"Cancer of the Urinary Bladder - Cancer Stat Facts." *SEER*, seer.cancer.gov/statfacts/html/urinb.html.

Results:

The results from Tests 1.1, 1.2, 1.3, 1.4 and 1.5 under PART A are summarized in the table below.

Table 1: Results for the accuracy of diagnosis of individual cancer-diagnosing tests

		Accuracy of Diagnosis	
Test Number	Test/cancer	In percentage	In decimals
1.1	Mammography/breast	54.1	0.541
1.2	Needle biopsy/lung	69.5	0.695
1.3	PSA/prostate	35.1	0.351
1.4	Ultrasonography/pancreatic	0.0114	0.000114
1.5	Urography/bladder	4.47	0.0447

PART B- MARKOV CHAINS

As mentioned on page 4, I used Markov Chains because it was the best fit for my investigation, despite the fact that I was required to go beyond my syllabus. The values in the matrices constructed for PART B were taken from the probability tree diagrams of the respective cancer tests from PART A.

Test 2.1- Accuracy of diagnosis of breast cancer using digital mammography screening (2015 study)

Initial state distribution matrix:

[patient has cancer   patient doesn't have cancer]

$$\begin{bmatrix} 0.125 & 0.875 \end{bmatrix}$$

Transition probability matrix:

$$\begin{bmatrix} \text{patient has cancer/positive test} & \text{patient has cancer/negative test} \\ \text{patient doesn't have cancer/positive test} & \text{patient doesn't have cancer/negative test} \end{bmatrix}$$

$$\begin{bmatrix} 0.999 & 0.00100 \\ 0.121 & 0.879 \end{bmatrix}$$

Probability of cancer occurrence in the following year of the study (2016) given false positive and false negative rates remain the same and test in question is used:

$$\begin{bmatrix} 0.125 & 0.875 \end{bmatrix} \begin{bmatrix} 0.999 & 0.00100 \\ 0.121 & 0.879 \end{bmatrix} = \begin{bmatrix} 0.231 & 0.769 \end{bmatrix}$$

Probability of cancer occurrence following two years of the original study (2017) given false positive and false negative rates remain the same and test in question is used:

$$\begin{bmatrix} 0.231 & 0.769 \end{bmatrix} \begin{bmatrix} 0.999 & 0.00100 \\ 0.121 & 0.879 \end{bmatrix} = \begin{bmatrix} 0.324 & 0.676 \end{bmatrix}$$

Test 2.2- Accuracy of diagnosis of lung cancer using CT-Guided Automated Needle Biopsy of Lung Nodules

Initial state distribution matrix:

[patient has cancer   patient doesn’t have cancer]

$[0.0627 \quad 0.937]$

Transition probability matrix:

$$\begin{bmatrix} \text{patient has cancer/positive test} & \text{patient has cancer/negative test} \\ \text{patient doesn't have cancer/positive test} & \text{patient doesn't have cancer/negative test} \end{bmatrix}$$

$$\begin{bmatrix} 0.829 & 0.171 \\ 0.0244 & 0.976 \end{bmatrix}$$

Probability of cancer occurrence in the following year of the study (2001) given false positive and false negative rates remain the same and test in question is used:

$$[0.0627 \quad 0.937]\begin{bmatrix} 0.829 & 0.171 \\ 0.0244 & 0.976 \end{bmatrix} = [0.0748 \quad 0.925]$$

Probability of cancer occurrence following two years of the study (2002) given false positive and false negative rates remain the same and test in question is used:

$$[0.0748 \quad 0.925]\begin{bmatrix} 0.829 & 0.171 \\ 0.0244 & 0.976 \end{bmatrix} = [0.0846 \quad 0.916]$$

Test 2.3- Accuracy of diagnosis of prostate cancer using prostate-specific antigen (PSA) testing and transrectal ultrasound-guided (TRUS) prostate biopsy

Initial state distribution matrix:

[patient has cancer   patient doesn’t have cancer]

$[0.111 \quad 0.889]$

Transition probability matrix:

$$\begin{bmatrix} \text{patient has cancer/positive test} & \text{patient has cancer/negative test} \\ \text{patient doesn't have cancer/positive test} & \text{patient doesn't have cancer/negative test} \end{bmatrix}$$

$$\begin{bmatrix} 0.650 & 0.350 \\ 0.150 & 0.850 \end{bmatrix}$$

Probability of cancer occurrence in the following year of the study (2018) given false positive and false negative rates remain the same and the test in question is used:

$$[0.111 \quad 0.889]\begin{bmatrix} 0.650 & 0.350 \\ 0.150 & 0.850 \end{bmatrix} = [0.206 \quad 0.795]$$

Probability of the accuracy of cancer occurrence following two years of the study (2019) given false positive and false negative rates remain the same and the test in question is used:

$$[0.206 \quad 0.795]\begin{bmatrix} 0.650 & 0.350 \\ 0.150 & 0.850 \end{bmatrix} = [0.253 \quad 0.747]$$

Test 2.4- Accuracy of the diagnosis of pancreatic cancer using ultrasonography

Initial state distribution matrix:

[patient has cancer   patient doesn't have cancer]

$[0.0000475 \quad 0.999]$

Transition probability matrix:

$$\begin{bmatrix} \text{patient has cancer/positive test} & \text{patient has cancer/negative test} \\ \text{patient doesn't have cancer/positive test} & \text{patient doesn't have cancer/negative test} \end{bmatrix}$$

$$\begin{bmatrix} 0.670 & 0.330 \\ 0.280 & 0.720 \end{bmatrix}$$

Probability of cancer occurrence in the following year of the study (1979) given false positive and false negative rates remain the same and only the test in question is used:

$$[0.0000475 \quad 0.999] \begin{bmatrix} 0.670 & 0.330 \\ 0.280 & 0.720 \end{bmatrix} = [0.280 \quad 0.719]$$

Probability of the accuracy of cancer occurrence following two years of the original study (1980) given false positive and false negative rates remain the same and the test in question is used:

$$[0.280 \quad 0.719] \begin{bmatrix} 0.670 & 0.330 \\ 0.280 & 0.720 \end{bmatrix} = [0.389 \quad 0.610]$$

Test 2.5- Accuracy of the diagnosis of bladder cancer using CT urography

Initial state distribution matrix:

[patient has cancer   patient doesn't have cancer]

$[0.0240 \quad 0.976]$

Transition probability matrix:

$$\begin{bmatrix} \text{patient has cancer/positive test} & \text{patient has cancer/negative test} \\ \text{patient doesn't have cancer/positive test} & \text{patient doesn't have cancer/negative test} \end{bmatrix}$$

$$\begin{bmatrix} 0.863 & 0.137 \\ 0.453 & 0.547 \end{bmatrix}$$

Probability of cancer occurrence in the following year of the study (2019) given false positive and false negative rates remain the same and only the test in question is used:

$$[0.0240 \quad 0.976]\begin{bmatrix} 0.863 & 0.137 \\ 0.453 & 0.547 \end{bmatrix} = [0.463 \quad 0.537]$$

Probability of the accuracy of cancer occurrence following two years of the original study (2020) given false positive and false negative rates remain the same and the test in question is used:

$$[0.463 \quad 0.537]\begin{bmatrix} 0.863 & 0.137 \\ 0.453 & 0.547 \end{bmatrix} = [0.643 \quad 0.357]$$

Results:

The results from Test 2.1, 2.2, 2.3, 2.4 and 2.5 conducted under PART B are summarized below.

Table 2: Probability of cancer occurrence following the years of the study from which false negative and false positive rates were taken, given false positive and false negative rates remain the same and test in question is used

Test Number	Type of Cancer	Year after the study	Probability (%)	Two years after the study	Probability (%)
2.1	Breast	2016	23.1	2017	32.4
2.2	Lung	2001	7.48	2002	8.46
2.3	Prostate	2018	20.6	2019	25.3
2.4	Pancreatic	1979	28.0	1980	38.9
2.5	Bladder	2019	46.3	2020	64.3

Conclusion:

The main conclusion that we can draw from PART A is that medical tests do not guarantee 100% accuracy when diagnosing a disease. Therefore, it is important for them to retake the test or try a different type of test to ensure that they certainly have or do not have the disease. For example, a mammography is not the only way breast cancer can be detected. The patient can opt to undergo a breast MRI scan or even a CT scan. From Table 1, the highest accuracy was given by the needle biopsy for detecting lung cancer at 69.5%, and lowest accuracy was given by the CT urography for detecting bladder cancer at 0.0447%.

The main conclusion drawn from PART B is that the probability of cancer occurrence has increased following the years of the study. We can observe from Table 2 that the probability of the occurrence of bladder cancer increased the most within a span of a year (increase in probability was 18% from 2019 to 2020), as compared to other types of cancers. The probability of the occurrence of lung cancer increased the least within a year (increase in probability was 0.98% from 2001 to 2002).

Limitations of the Exploration:

There are a few limitations of PART B of the investigation. Firstly, the probability has too many conditions. Participants would have to take the exact same test as mentioned in the study, which is highly unlikely given that there are many types of medical tests that can detect cancer. Secondly, the statistics for cancer occurrence and the false negative/false positive rates were not taken from the same year. For example, the probability of a female having breast cancer was based on a 2020 statistic, but the false negative/false positive rates were taken from a study published in 2015. Next, the false positive/false negative rates for an ultrasonography test to detect pancreatic cancer was taken from a relatively old study. This implies that technology might have advanced during the current years, thus the findings of Test 2.4 under PART B may not be applicable to present day.

References:

1. “Cancer of the Urinary Bladder - Cancer Stat Facts.” *SEER*, seer.cancer.gov/statfacts/html/urinb.html.

2. Fitzgerald, P J, et al. “The Value of Diagnostic Aids in Detecting Pancreas Cancer.” *Cancer*, U.S. National Library of Medicine, Mar. 1978, pubmed.ncbi.nlm.nih.gov/638974/.

3. “Key Statistics for Prostate Cancer: Prostate Cancer Facts.” *American Cancer Society*, www.cancer.org/cancer/prostate-cancer/about/key-statistics.html.

4. Krasnow, Ross E., and Lambros Stamatakis. “What to Do When Prostate Cancer Biopsy/PSA Test Results Conflict.” *MedStar Washington Hospital Center Blog Center View*, 28 Sept. 2019, blog.medstarwashington.org/2017/09/28/prostate-cancer-high-psa-negative-biopsy/.

5. “Lung Cancer Statistics: How Common Is Lung Cancer.” *American Cancer Society*, www.cancer.org/cancer/lung-cancer/about/key-statistics.html.

6. Nelson, Heidi D, et al. “Factors Associated With Rates of False-Positive and False-Negative Results From Digital Mammography Screening: An Analysis of Registry Data.” *Annals of Internal Medicine*, U.S. National Library of Medicine, 16 Feb. 2016, www.ncbi.nlm.nih.gov/pmc/articles/PMC5091936/.

7. Rawla, Prashanth, et al. “Epidemiology of Pancreatic Cancer: Global Trends, Etiology and Risk Factors.” *World Journal of Oncology*, Elmer Press, Feb. 2019, www.ncbi.nlm.nih.gov/pmc/articles/PMC6396775/.

8. Trinh , Tony W, et al. “Bladder Cancer Diagnosis with CT Urography: Test Characteristics and Reasons for False-Positive and False-Negative Results.” *Abdominal Radiology (New York)*, U.S. National Library of Medicine, Mar. 2018, pubmed.ncbi.nlm.nih.gov/28677000/.

9. Tsudaka, Hiroshi, et al. *Diagnostic Accuracy of CT-Guided Automated Needle Biopsy of Lung Nodules*. July 2000, www.ajronline.org/doi/pdf/10.2214/ajr.175.1.1750239.

10. “U.S. Breast Cancer Statistics.” *Breastcancer.org*, 4 Feb. 2021, www.breastcancer.org/symptoms/understand_bc/statistics.