



Retrospective Analysis of Hematological Cancer

By

Nurbinta Sultana

ID: 161-35-1444

Supervisor: Mr. Khaled Sohel

Assistant Professor

A thesis submitted in partial fulfillment of the requirement for the degree of Bachelor
of science in software engineering

Department of Software Engineering
Daffodil International University

FALL 2019

APPROVAL

This Thesis titled “**Retrospective Analysis of Hematological Cancer**”, submitted by **Nurbinta Sultana, ID No: 161-35-1444** to the Department of software engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillments for the requirements of the degree of B.Sc. in Software Engineering and approved as its style and contents.

BOARD OF EXAMINERS

Chairman

Dr. Touhid Bhuiyan
Professor and Head
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 1

Dr. Md. Asraf Ali
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner2

Asif Khan Shakir
Lecturer
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

External Examiner

Prof. Dr. Mohammad Abul Kashem
Professor
Department of Computer Science and Engineering
Faculty of Electrical and Electronic Engineering
Dhaka University of Engineering & Technology, Gazipur

Daffodil International University

THESIS DECLARATION

The thesis entitled “**Retrospective Analysis of Hematological Cancer**” is done under the supervision of M. Khaled Sohel, Assistant Professor, Department of Software Engineering, Daffodil International University. I declare that this thesis is my original work for the degree of B.Sc. in Software Engineering and that neither the whole work nor any part has been submitted for another degree in this or any other university.

Submitted by 

Name: Nurbinta sultana

ID: 161-35-1444

Batch: 19th

Department of Software Engineering

Faculty of Science and Information Technology

Daffodil International University

Certified by:



M. Khaled Sohel

Assistant Professor

Department of Software Engineering

Faculty of Science and Information Technology

Daffodil International University

ACKNOWLEDGEMENT

I am thankful to my creator for giving me the opportunity to complete this research work and learn so much. I am thankful to my research supervisor, M. Khaled Sohel, for providing careful guidance starting from selecting the research scope to successfully finalizing the research work. I also thank Dr. Md. Asraf Ali for his valuable observation, which was always insightful. Finally, I want like to express my gratitude to Professor Dr. Touhid Bhuiyan, head of the Software engineering faculty, for inspiring us in all means.

TABLE OF CONTENTS

APPROVAL	II
THESIS DECLARATION	Error! Bookmark not defined.
ACKNOWLEDGEMENT	IV
TABLE OF CONTENTS	V-VII
LIST OF TABLE	VIII-IX
LIST OF CHART	X
LIST OF ABBREVIATIONS	XI
ABSTRACT	XII
CHAPTER 1: INTRODUCTION	1-5
1.1 Background	1
1.1.1 Scope and severity	1
1.2 Motivation	2
1.3 Problem statement	3
1.4 Research questions	4
1.5 Objectives	4
1.6 Research scope	4
1.7 Thesis organization	5
CHAPTER 2: LITERATURE REVIEW	6-8
2.1 Previous research on hematological cancer	6
2.2 Research gap	8
2.3 Summary	8
CHAPTER 3: RESEARCH METHODOLOGY	9-11
3.1 Methodology	9
3.2 Methodology Flow chart	10

3.3 Data description	11
3.4 Data preprocessing	11
3.5 Data analysis	11
3.6 Summary	11
CHAPTER 4: RESULTS AND DISCUSSION	13-34
4.1 Results	13
4.2 Data visualization according to AGE	13
4.2.2 Hypothesis testing	14
4.3 Data visualization according to GENDER	15
4.3.2 Hypothesis testing	15
4.4 Data visualization according to patient's occupation	17
4.4.2 Hypothesis testing	18
4.5.2 Hypothesis Testing	21
4.6 Data visualization according to previous cancer history in relatives	22
4.6.2 Hypothesis Testing	22
4.7 Data visualization according to Presence or absence of associated disease or infection.	24
4.7.2 Hypothesis Testing	25
4.8 Data visualization according to have proper idea about the cause of cancer.	26
4.8.2 Hypothesis testing	27
4.9 Data visualization according to diagnosed cancer	29
4.10 Multiple Linear Regression Model	29
4.10.4 Hypothesis testing	33
4.11 Discussion	34
CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS	37-36
5.1 Findings & Contribution	37

5.2 Limitation	37
5.3 Recommendations future works	38
References:	39-38

LIST OF TABLE

Table 4.2.1: Distribution of hematological malignancy according to age	13
Table 4.2.3: Determination of age related hypothesis by ANNOVA Test	14
Table 4.3.1: Distribution of the patients of hematological malignancies according to Gender	15
Table 4.3.3: Determination of gender related hypothesis by Independent sample T-test (part 1,part 2, part 3)	16-17
Table 4.4.1: Distribution of the patients of hematological malignancies according to their occupation.	17
Table 4.4.3: Determination of occupation related hypothesis by One way ANNOVA Test	18
Table 4.5.1: Distribution of the patients of hematological malignancies according to their residence.	19
Table 4.5.3: Determination of residence related hypothesis by One way ANNOVA Test	20
Table 4.6.1: Distribution of the patients of hematological malignancies of Bangladesh according to previous cancer history in relatives.	21
Table 4.6.3: Determination of hypothesis related previous cancer history in relatives by Independent samples test (part 1,part 2)	21-22
Table 4.6.4: T-test for Equality of Means	22
Table 4.7.1: Distribution of the patients of hematological malignancies according to Presence or absence of associated disease or infection.	23
Table 4.7.3: Determination of hypothesis related to associated disease or infection by Independent samples test (part 1, part 2)	24

Table 4.7.4: T-test for Equality of Means	24
Table 4.8.1: Distribution of the patients of hematological malignancies by having proper idea about the cause of cancer.	25
Table 4.8.3: Determination of hypothesis related to Idea about the cause of cancer by Independent samples T-test.(part 1, part 2, part 3)	26-27
Table 4.9.1: Distribution of the patients of hematological malignancies according to the diagnosed cancer	27-28
Table 4.10.1 Multiple Linear Regression Model	28
Table 4.10.2 Model Summary	29
Table 4.10.3: Coefficients	29
Table 4.10.5: Determination of Hypothesis related to the regression model by ANNOVA test	31

LIST OF CHART

Charts1: Observed Cum Prob Scatterplot	30
Chart 2: Regression Standardized Predicted Value	31

LIST OF ABBREVIATIONS

Abbreviation	Explanation
SPSS	Statistical Package for the Social Sciences
HM	Hematological Malignancy
AML	Acute myeloid leukemia
ALL	Acute lymphoblastic leukemia
CML	Chronic myeloid leukemia
CLL	Chronic Lymphocytic leukemia
HL	Hodgkin's lymphoma
NHL	Non-Hodgkin's lymphoma

ABSTRACT

Hematological Cancer incidence is increasing day by day especially in low income countries which become a global burden now a days. As Bangladesh is a low-income country, cancer incidence observe rapid aging here. Besides cancer treatment is very costly & no proper antidote is available. So preventive measure is the best way to stay safe from hematological cancer. To know the preventive measures, we have to find out the risk factor.

Our objective is to find out the risk factor & asses the current Hematological Malignancy status of this country and to predict further Hematological Malignancy incidence for awareness.

A retrospective descriptive type of study was done over 505 patients of hematological cancer of 2019. Data were collected from patient through face to face interview. SPSS tool, Hypothesis testing & Multiple linear regression model are used to analyze data. We correlate Hematological malignancy with age, gender, occupation, residence, associated infection, idea about cause of cancer and previous cancer history in relative.

A total number of 505 patient aged from 4 to 92 years who had been suffering from hematological malignancy. The study show that the most frequent leukemia was acute lymphoblastic leukemia about 31.7%, and the least was chronic lymphocytic leukemia about 6.3%. By hypothesis testing we find that Age, Gender, Occupation & idea about the cause of cancer has impact on Hematological cancer. Hematological Cancer can occur at any age but the most vulnerable age group is 10-19 (28.3%) & 2nd most is 20-29 (19.6%). In case of Gender males were 63.6% (n = 321) and females were 36.4% (n = 184) and male - female ratio is 1.7:1 which means that male has more risk at Hematological cancer. In occupation, our study show that Students are the most vulnerable group about 44% and the least vulnerable group is Housewives which is about 8.3%. The reason behind this may be radiation from electric device such as Mobile & Laptops which are mostly use by the Student. We also find that most of the patient about 80.6% has no proper idea about the cause of cancer. Without knowing the proper

cause of any incidence, one can never prevent it. So, it could be the important factor in Hematological cancer occurrences.

Our study shows the risk factor and the distribution pattern of hematological malignancy in the area of Bangladesh. It presents the distribution pattern of hematological cancer according to Age, Gender & correlation of Hematological Malignancy with occupation, residence & other factor.

CHAPTER 1

INTRODUCTION

1.1 Background

In western country patient's information are kept properly by the hospital and health care providers. This is very helpful and stander for collection data. In our country some specialized hospital follow this system properly. Moreover there are few cancer specialized hospital in this country [1]. So proper data collection from hospital record is difficult here. Measurement of cancer patients and proper diagnosis of cancer are the essential key attributes in the process. Most of the tertiary health care facilities for the treatment of cancer are available in the capital city where about 70% of total population live in village [16]. As a result patients who are seeking for the treatment for the cancer have to travel Dhaka. Many poor people cannot afford the total cost of treatment and travelling. As a result they remain untreated and their cases remain unknown. In Bangladesh there is no population based or national cancer registry. So the cancer status is largely unknown in this country [1]. Predicting current cancer status and the risk factor of cancer occurrence are the objective of this research where SPSS tools is applied and hypothesis testing is used to find a comparative analysis between cancer and different risk factor.

1.1.1 Scope and severity

Hospitals and nursing home of developed countries like USA and UK apply a different rating system of cancer incidence .They periodically measures the cancer incidence and prevalence. The term incidence refers the number of new patients in that period or year. The term prevalence

refers the number of new and previous cases up to that period or year. But this type of service system is unavailable in our country.

Cancer is now a growing problem in all over the world. Besides it has no proper antidote or vaccine and no affordable treatment. The treatment of cancer based on chemotherapy is very costly. As a result the rate of death in cancer is very high. Cancer not only cause a death but also destroy a family socially and economically. A cancer patient become a burden to his/her family. So several researches are needed to find out an affordable treatment policy and vaccine to prevent cancer disease. World Health Organization (WHO) has already declared some preventive measures. They are Tobacco control, Overweight & obesity prevention & control, Physical inactivity prevention & control, Harmful use of alcohol prevention & Control, National Immunization [9]. Now we need to find out cost effective treatment policy & vaccine. There are various types of cancer such as Mouth and oropharynx cancer; esophagus cancer; Trachea, Bronchus & Lung Cancer; Stomach Cancer, Colorectal cancer and Hematological Cancer. In Bangladesh more than 1, 00,000cancer death in 2018 accounting for 10%of all deaths in the country [10]. New cases are predicted to rise from 150,781 cases in 2018 to 250,726 cases in 2035 [10]. In that case cancer will be the great cause of death in near future in our country. Besides it can harm the whole country's socio-economic condition. Bangladesh is now a developing country but it may be a great obstacle in the way of becoming a develop country. So we need more focus on this tropics.

1.2 Motivation

Hematologic malignancies are forms of cancer that begin in the cells of blood-forming tissue, such as the bone marrow, or in the cells of the immune system. There are different type of

hematological malignancies such as Acute Lymphoblastic Leukemia, Acute Meroblastic leukemia, Chronic Lymphocytic Leukemia, Chronic Myeloid Leukemia, Hodgkin's Lymphoma, Non-Hodgkin's Lymphoma etc. When we start our study about Hematological malignancies, we found some article & journal about the topic. According to **WHO**, Bangladesh is increasing cancer incidence with 122716 new cases of cancer in 2012 [11]. In 2030, these will be increase about 77% [11]. In previous study, we see the pattern and distribution of hematological cancer in Bangladesh [1]. We also see that hematological malignancies are more common in developed & developing country. Ours is a developing country & cancer patients increasing in our country day by day. Besides there is no proper treatment of cancer & the cancer treatment is very costly. In that case we need to emphasis on preventive measures by raising awareness.

To find out the preventive measures, we have to find out the risk factor. In this research, our main objective is trying to find out the risk factors. So we choose the topic.

Besides there is no proper treatment of cancer & the cancer treatment is very costly. In that case we need to emphasis on preventive measures by raising awareness.

1.3 Problem statement

Hematological cancer is now a growing problem in all over the world. Besides it has no proper antidote or vaccine and no affordable treatment. The treatment of hematological cancer based on chemotherapy is very costly. As a result the rate of death in hematological cancer is very high. Hematological Cancer not only cause a death but also destroy a family socially and economically. A hematological cancer patient become a burden to his/her family. So several researches are needed to find out preventive measures and affordable treatment policy of hematological cancer.

To find out the preventive measures we have to find out the risk factor. To find out the risk factor we have to know the pattern and distribution HM and current status of HM in Bangladesh. World Health Organization (WHO) has already declared some preventive measures for cancer. They are Tobacco control, Overweight & obesity prevention & control, Physical inactivity prevention & control, Harmful use of alcohol prevention & Control, National Immunization [9]. We will try to find some risk factor for HM.

1.4 Research questions

The research questions are following to determine efficiency among different dataset combinations:

1. What is the current situation of Hematological malignancy in Bangladesh?
2. What are the most responsive risk factors of Hematological malignancy in Bangladesh?

1.5 Objectives

1. To assess the current hematological incidence in Bangladesh.
2. To find out the risk factor of hematological malignancy in Bangladesh.

1.6 Research scope

We had identified few research gaps and based on these lacking we select Statistical Package for the Social Sciences (SPSS) tool to predict the risk factor the hematological cancer. After applying

this tools we get different data set in testing phase. Then we use hypothesis testing such as **T-test**, **ANNOVA**-test for analyzing the effectiveness of this prediction. Then finally, we compared among these different predictions to find an effective combination.

1.7 Thesis organization

In the following chapters, we discussed the literature including the research gap, described the research methodology along with data description, mentioned the results and discussions and finally discussed the conclusion with recommendations with our research findings, limitations and directions.

CHAPTER 2

LITERATURE REVIEW

2.1 Previous research on hematological cancer

Cancer is becoming a leading cause of death and disability in low-income countries [1]. The global cancer burden is estimated to have risen to 18.1 million new cases and 9.6 million death in 2018. One in 5 men and one in 6 women worldwide develop cancer during their life time, and one in 8 men and one in 11 women die from this disease. The total number of people who are alive with in five year of cancer diagnosis is called 5-year prevalence, is estimated to be 43.8 million [2]. It is projected that by 2030 between 10 and 11 million cancer will be diagnosed each year in low and middle-income countries and death from cancer worldwide reach over 13 million in 2030 [3]. Despite current or future efforts, the challenges of tackling cancer are enormous, since the occurrence of cancer in these setting will continue to rise due to increasing lifespans through better control of communicable disease [1]. Bangladesh shows major advances in relation to the management of infectious diseases as recently highlighted in Lancet [4]. The chronic disease in particular cancer are less prioritized [5]. The status of cancer in this country is largely unknown, as there is no population-based cancer registry nor a national cancer registry of any other kind. At present there are about 13 to 15 lakh cancer patients in Bangladesh, with about two lakh patients newly diagnosed with cancer each year [6]. According to the Bangladesh Bureau of statistics, cancer is the sixth leading cause of death. International agency for research on cancer has estimated cancer related death rates in Bangladesh to be 7.5% in 2005 and rises to 13% in 2030 [5]. The real cancer status may not be reflected by these estimate as many cases go unreported because of lacking of education, poverty, misconception and awareness among all

population. Moreover, the cost of cancer treatment is very high so that the cancer care and management systems in Bangladesh are below international standard level. Besides, there are lack of oncologists and sufficient infrastructure in Bangladesh.

Hematological malignancies are primary cancers originating cells of the bone marrow and lymphatic system. There are three main categories which are Leukemia, Lymphoma, and multiple myeloma (MM). Other categories are Myelodysplastic syndrome (MDS), polycythemia Vera, and primary myelofibrosis. The most common type of leukemia are: Chronic myeloid leukemia (CML), Chronic lymphocytic leukemia (CLL), Acute myeloid leukemia (AML), Acute lymphoblastic leukemia (ALL). Lymphomas are two types. They are Hodgkin's Lymphoma (HL) & Non-Hodgkin's Lymphoma (NHL). The Primary causes of hematological malignancy are remain unknown but it is believed to be connected with environmental exposures, ionizing radiation and infectious agent. These includes: x-ray, pesticides, Dye, Benzene vapor, cytotoxic drugs and chromosomal [1]. There is a link in between infectious agent and hematological malignancies in the young. Epstein-bar virus (EBV) is the major cause of Burkett's lymphoma and Hodgkin's disease. Human T cell leukemia virus, herpes virus 6, 7 and 8 also cause different types of hematological malignancy [7]. Information pertaining to the epidemiological aspect of Hematological Malignancy in Bangladeshi population is limited [8]. To identify the risk factor of hematological malignancy, we have to find out the epidemiological aspect of hematological malignancy in the environmental background. It would provide epidemiological aspect of hematological basis for devising the cancer care management and preventive strategies of these malignancies. In the present retrospective study, we report the overall pattern and age distribution of hematological malignancies in Bangladesh.

2.2 Research gap

Here are few research gaps that I have found and focusing on these lacking, I have conducted this research:

Review of literature suggests that a few researches have been carried out to study the hematological malignancies.

There is a need for better understanding of the data attributes which are essential to describe the data structure.

2.3 Summary

This research is motivated by the need for concentrating study to predict the risk factor of hematological cancer with the help of data analysis by using SPSS tool and hypothesis testing.

CHAPTER 3

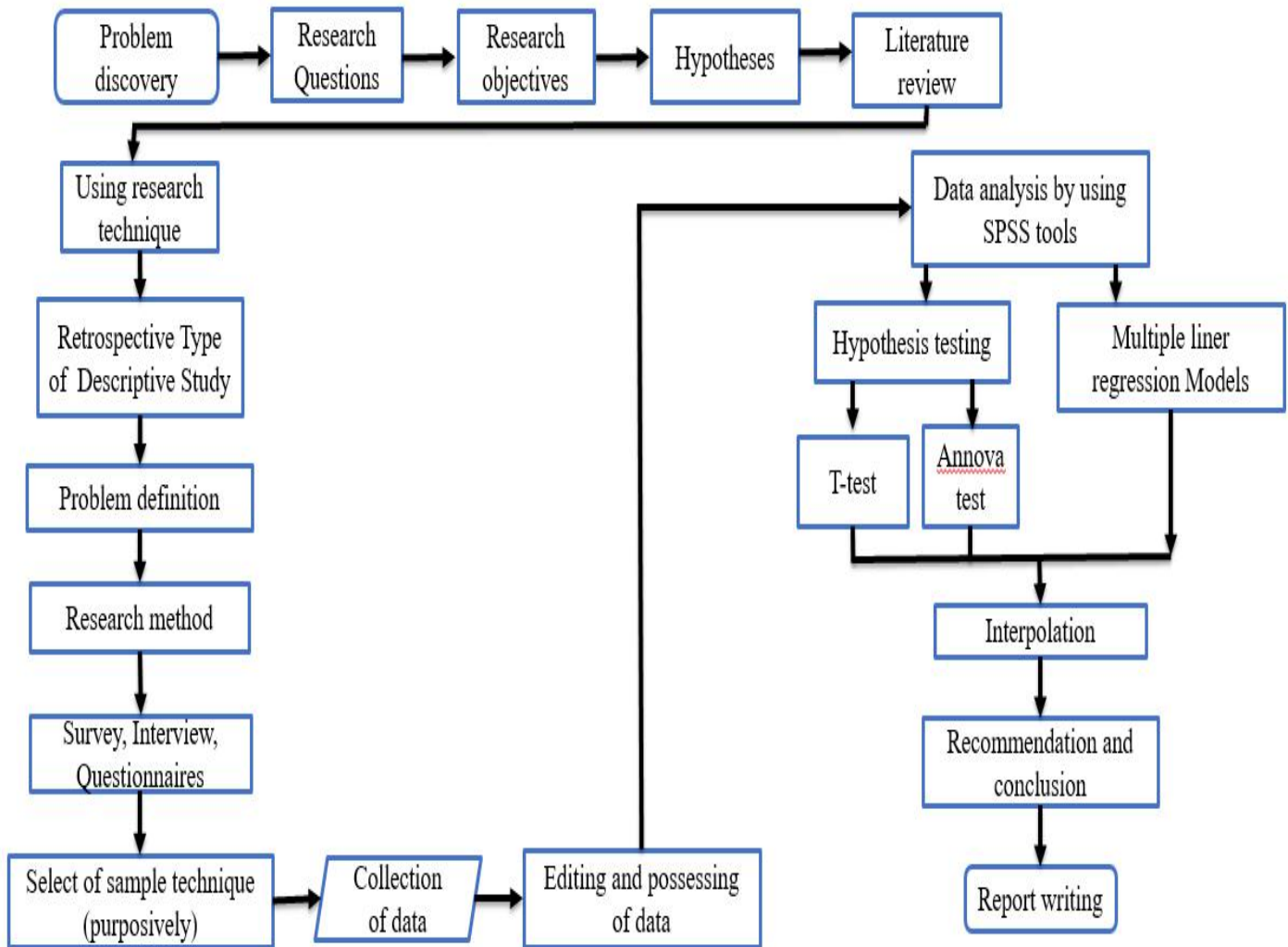
RESEARCH METHODOLOGY

3.1 Methodology

For research methodology, we have used Statistical Package for the Social Sciences (SPSS) to predict the risk factor of hematology cancer. It was a retrospective descriptive study and our study area was division based. Man and women of all age group was our study population. In Bangladesh, there is no proper and effective hospital recode keeping system. As a result there are possibilities that same patient may visit multiple time in the same hospital or different hospital. To avoid duplicate, reenrolled cases, we collect data from the patient by face to face interview and analysis the data by using SPSS tool.

Finally we make combination of analysis data to get a better prediction by using **Multiple liner regression model**.

3.2 Methodology Flow chart



3.3 Data description

Data are collected from diagnosed hematological cancer patients. We visited Dhaka medical college several time and talk with patients and gather information. Thus we collect row data for our research.

3.4 Data preprocessing

After collecting row data by face to face interview from the patients we process these data according to our research question. We select about 500 patient's data for our research work. We exclude some data due to lack of proper information, this similarity among information's.

3.5 Data analysis

After collecting research data we made a master table by using Excel. Then we converted the data from Excel to SPSS tool. By Using SPSS tool, we separated the data chronologically in different groups. By hypothesis testing such as T-test, ANNOVA-test was done among these groups. Finally we used multiple linear regression model to find out the comparison among these groups.

3.6 Summary

It was a Retrospective Descriptive study and Study area was all over Bangladesh. About 505 patients data were collected which include men and women of all age group. : Mixed

(Qualitative & Quantitative) types data are presented here. Data were collected by face to face interview through questioner.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Results

The study found following results through data collected by hospital records which are noted below accordingly.

4.2 Data visualization according to AGE

Table 4.2.1: Distribution of hematological malignancy according to age.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	under 9	8	1.6	1.6	1.6
	10-19	143	28.3	28.3	29.9
	20-29	99	19.6	19.6	49.5
	30-39	91	18.0	18.0	67.5
	40-49	62	12.3	12.3	79.8
	50-59	53	10.5	10.5	90.3
	60-69	42	8.3	8.3	98.6
	70 and older	7	1.4	1.4	100.0
	Total	505	100.0	100.0	

The table shows that most of the patient of hematological malignancy are is the age group of 10-19 years old and 2nd most group is 20-29 years old.

4.2.2 Hypothesis testing

Null Hypothesis: Age has no impact on cancer incidence.

Alternative Hypothesis: Age has impact on cancer incidence.

Table 4.2.3: Determination of age related hypothesis by ANNOVA Test.

Diagnosed Cancer	Sig.
Between Groups (Combined)	.000
Linear Term	
Unweighted	.319
Weighted	.000
Deviation	.000
Within Groups	
Total	

The test is done to know which hypothesis is statistically important and it could measure by the significance value.

Result for Age: Here we see that the significance level is .000 which is less than .05. So the null hypothesis is rejected and alternative hypothesis is accepted. As a result we can say that Age has impact on Cancer.

4.3 Data visualization according to GENDER

Table 4.3.1: Distribution of the patients of hematological malignancies according to Gender.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	321	63.6	63.6	63.6
	Female	184	36.4	36.4	100.0
	Total	505	100.0	100.0	

The table shows that most of the patients (63.6%) of hematological malignancies are male.

4.3.2 Hypothesis testing

Null hypothesis: Gender has no impact on Hematological malignancy

Alternative Hypothesis: Gender has impact on Hematological malignancy.

Table 4.3.3: Determination of gender related hypothesis by Independent sample T-test.

Part: 1

	gender	N	Mean	Std. Deviation	Std. Error Mean
Diagnosed Cancer	Male	321	1.94	1.370	.076
	Female	184	2.27	1.426	.105

Here we see that there is large difference between mean & standard deviation which means that there is little or no chance of equal variances.

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
Diagnosed Cancer	Equal variances assumed	.560	.455	-2.507	503
	Equal variances not assumed			-2.480	368.995

Part: 2

Here significance value is .455 which is greater than .05 that means in case of 2 tailed we take the significance level of equal variance section.

		Sig. (2-tailed)	Mean Difference	Std. Error Difference
Diagnosed Cancer	Equal variances assumed	.012	-.322	.129
	Equal variances not assumed	.014	-.322	.130

Part: 3

Here the significance value is .012 which is less than .05.

The test is done to know which hypothesis is statistically important. It could measure by the significance value which is .012.

Result for GENDER: Here the significance level (2-tailed) is .012 which is less than .05. It means that null hypothesis is rejected. So we can say that gender has impact on hematological malignancy.

4.4 Data visualization according to patient's occupation

Table 4.4.1: Distribution of the patients of hematological malignancies according to their occupation.

		Frequency	Percent	Valid Percent

Valid	Student	222	44.0	44.0
	jobholder	154	30.5	30.5
	jobless	35	6.9	6.9
	business	48	9.5	9.5
	housewife	42	8.3	8.3
	baby	4	.8	.8
	Total	505	100.0	100.0

The table shows that most of the respondents (44%) were student which means that Students are more vulnerable to Hematological Malignancy.

4.4.2 Hypothesis testing

Null Hypothesis: Occupation has no impact on Hematological Malignancy.

Alternative Hypothesis: Occupation has impact on Hematological Malignancy.

Table 4.4.3: Determination of occupation related hypothesis by One way ANNOVA Test.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	39.555	5	7.911	4.175	.001
Within Groups	945.542	499	1.895		
Total	985.097	504			

The test is done to know which hypothesis is statistically important. It could measure by the significance value which is .001.

Result for patient’s occupation: Here the significance level is .001 which is lesser than 0.05. It means that the null hypothesis is rejected. So alternative hypothesis is accepted here which means that at least one pair of occupation category has different mean on hematological malignancy. As a result we can say occupation has impact on Hematological Malignancy.

4.5 Data visualization according to patient’s residence

		Frequency	Percent	Valid Percent
Valid	Dhaka	178	35.2	35.2
	Chittagong	95	18.8	18.8
	Rangpur	60	11.9	11.9
	Barisal	11	2.2	2.2
	Rajshahi	49	9.7	9.7
	Mymensingh	46	9.1	9.1
	Sylhet	28	5.5	5.5
	Khulna	38	7.5	7.5

Table

	Total	505	100.0	100.0
--	-------	-----	-------	-------

 4.5.1: Distribution of the patients of hematological malignancies according to their residence.

The table shows that most of the patients of hematological malignancies of Bangladesh were from Dhaka and 2nd most from Chittagong.

4.5.2 Hypothesis Testing

Null Hypothesis: Residence has impact no impact on Hematological Malignancy.

Alternative Hypothesis: Residence has impact on Hematological Malignancy.

Table 4.5.3: Determination of residence related hypothesis by One way ANNOVA Test.

Diagnosed Cancer	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	17.582	7	2.512	1.290	.253
Within Groups	967.515	497	1.947		
Total	985.097	504			

The test is done to know which hypothesis is statistically important. It could measure by the significance value which is .253.

Result for patient's residence: Here the significance level is .253 which is greater than .05. It means that Null Hypothesis is accepted. It also means that there is no mean difference of residence on hematological Malignancy. As a result we can say that Residence has no impact on Hematological Malignancy.

4.6 Data visualization according to previous cancer history in relatives

Table 4.6.1: Distribution of the patients of hematological malignancies of Bangladesh according to previous cancer history in relatives.

		Frequency	Percent	Valid Percent
Valid	Absent	432	85.5	85.5
	Present	73	14.5	14.5
	Total	505	100.0	100.0

The table shows that most of the patients (85.5%) of hematological malignancies have no previous cancer history in relatives.

4.6.2 Hypothesis Testing

Null Hypothesis: Previous cancer history in relatives has no impact on Hematological malignancy.

Alternative Hypothesis: Previous cancer history in relatives has impact on Hematological malignancy.

Table 4.6.3: Determination of hypothesis related previous cancer history in relatives by Independent samples test.

	Previous cancer history in relatives	N	Mean	Std. Deviation	Std. Error Mean
Diagnosed Cancer	Absent	432	2.09	1.410	.068
	Present	73	1.90	1.325	.155

Part: 1

Here we see that there is large difference between mean & standard deviation which means that there is little or no chance of equal variances.

Part: 2

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.1	t	df
Diagnosed Cancer	Equal variances assumed	3.402	.066	1.039	503
	Equal variances not assumed			1.086	101.582

Here significance value is .066 which is greater than .05 that means in case of 2 tailed we take the significance level of equal variance section.

		Sig. (2-tailed)	Mean Difference	Std. Error Difference
Diagnosed Cancer	Equal variances assumed	.299	.184	.177
	Equal variances not assumed	.280	.184	.169

Table 4.6.4: T-test for Equality of Means

Here the significance value is .299 which is greater than .05. The test is done to know which hypothesis is statistically important. It could measure by the significance value which is .299.

Result according to previous cancer history in relatives:

Here the significance level is .299, which is greater than .05. So the null hypothesis is accepted and alternative hypothesis is rejected which means that there is no impact of previous cancer history in relatives on Hematological Malignancy.

4.7 Data visualization according to Presence or absence of associated disease or infection.

		Frequency	Percent	Valid Percent
Valid	Absent	445	88.1	88.1
	Present	60	11.9	11.9
	Total	505	100.0	100.0

Table
4.7.1:
Distrib
ution
of the

patients of hematological malignancies according to Presence or absence of associated disease or infection.

The table shows most of the patients (88.1%) of hematological malignancies have no associated disease or infection.

4.7.2 Hypothesis Testing

Null Hypothesis: Associated disease or infection has no impact on Hematological Malignancy.

Alternative Hypothesis: Associated disease or infection has impact on Hematological Malignancy.

Table 4.7.3: Determination of hypothesis related to associated disease or infection by Independent samples test.

Part: 1

	Associated disease	N	Mean	Std. Deviation	Std. Error Mean
Diagnosed Cancer	Absent	445	2.06	1.402	.066
	Present	60	2.05	1.383	.179

Her

e we see that there is large difference between mean & standard deviation which means that there is little or no chance of equal variances.

Part: 2

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
Diagnosed Cancer	Equal variances assumed	.332	.565	.067	503
	Equal variances not assumed			.068	76.278

Her

significance value is .565 which is greater than .05 that means in case of 2 tailed we take the significance level of equal variance section.

Table 4.7.4: T-test for Equality of Means

		Sig. (2-tailed)	Mean Difference	Std. Error Difference
Diagnosed Cancer	Equal variances assumed	.946	.013	.192
	Equal variances not assumed	.946	.013	.191

Here the significance value is .946 which is greater than .05.

The test is done to know which hypothesis is statistically important. It could measure by the significance value which is .946.

Result according to Presence or absence of associated disease or infection:

Here the significance level is .946 which is greater than .05. So the null hypothesis is accepted and alternative hypothesis is rejected which means that associated disease or infection has no impact on Hematological malignancy.

4.8 Data visualization according to have proper idea about the cause of cancer.

Table 4.8.1: Distribution of the patients of hematological malignancies by having proper idea about the cause of cancer.

		Frequency	Percent	Valid Percent
Valid	No	407	80.6	80.6
	Yes	98	19.4	19.4
	Total	505	100.0	100.0

The table shows most of the patients (80.6%) of hematological malignancies have no proper idea about the cause of cancer.

4.8.2 Hypothesis testing

Null Hypothesis: Idea about the cause of cancer has no impact on Hematological Malignancy.

Alternative Hypothesis: Idea about the cause of cancer has impact on Hematological Malignancy.

Table 4.8.3: Determination of hypothesis related to Idea about the cause of cancer by Independent samples T-test.

	Idea about Cause of Cancer	N	Mean	Std. Deviation	Std. Error Mean
Diagnosed Cancer	No	407	1.98	1.396	.069
	Yes	98	2.41	1.361	.137

Part: 1

Here we see that there is large difference between mean & standard deviation which means that there is little or no chance of equal variances.

Part: 2

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
Diagnosed Cancer	Equal variances assumed	.003	.953	-2.753	503
	Equal variances not assumed			-2.796	150.041

Here significance value is .953 which is greater than .05 that means in case of 2 tailed we take the significance level of equal variance section.

Part: 3

		Sig. (2-tailed)	Mean Difference	Std. Error Difference
Diagnosed Cancer	Equal variances assumed	.006	-.430	.156
	Equal variances not assumed	.006	-.430	.154

Here the significance value is .946 which is greater than .006.

The test is done to know which hypothesis is statistically important. It could measure by the significance value which is .006.

Result for having proper idea about the cause of cancer: Here the significance level (2-tailed) is .006 which is lesser than .05. So the null hypothesis is rejected and alternative hypothesis is accepted which means that Idea about cause of cancer has different mean value on

Hematological Malignancy. As a result, we can say that Idea about cause of cancer has impact on Hematological Malignancy.

4.9 Data visualization according to diagnosed cancer

Table 4.9.1: Distribution of the patients of hematological malignancies according to the diagnosed cancer.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NHL	56	11.1	11.1	11.1
	ALL	160	31.7	31.7	42.8
	AML	104	20.6	20.6	63.4
	CML	99	19.6	19.6	83.0
	HL	54	10.7	10.7	93.7
	CLL	32	6.3	6.3	100.0
	Total	505	100.0	100.0	

The table shows most of the patients (31.7%) of hematological malignancies are Acute Lymphoblastic Leukemia (ALL).

4.10 Multiple Linear Regression Model

Table 4.10.1: Multiple Linear Regression Model

	Mean	Std. Deviation	N
--	------	----------------	---

Diagnosed Cancer	2.06	1.398	505
gender	.36	.482	505
Occupation	1.10	1.326	505
Division	2.19	2.360	505
Previous cancer history in relatives	.14	.352	505
Associated disease	.12	.324	505
Idea about Cause of Cancer	.19	.396	505

Table 4.10.2: Model summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	Change Statistics	
						F Change	df1
1	.197 ^a	.039	.027	1.379	.039	3.361	6

. Predictors:

(Constant), Idea about Cause of Cancer, gender, previous cancer history in relatives, Division, Occupation, Associated disease.

b. Dependent Variable: Diagnosed Cancer.

Here we see that the R Square value is .039 which explain the variance between dependent and independent variable is 3.9%.

Table 4.10.3: Coefficients

		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.797	.117		15.311	.000
	gender	.308	.129	.106	2.392	.017
	Occupation	.092	.047	.087	1.949	.052
	Division	-.005	.026	-.008	-.180	.858
	Previous cancer history in relatives	-.164	.182	-.041	-.906	.365
	Associated disease	-.016	.195	-.004	-.080	.936
	Idea about Cause of Cancer	.449	.157	.127	2.857	.004

Regression Equation

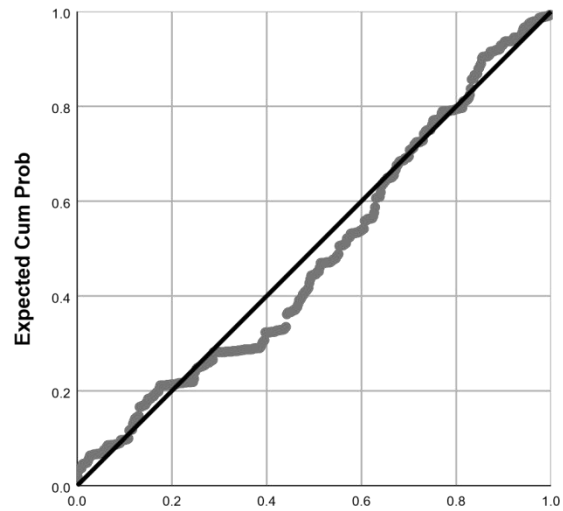
$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5.$$

Diagnosed Cancer = 1.797 + .308 (gender) + .092(occupation) - .005(Division) - .164(p.c.h.) - .016(a. disease) + .449 (idea about cause of cancer)

Chart 1: Observed Cum Prob Scatterplot

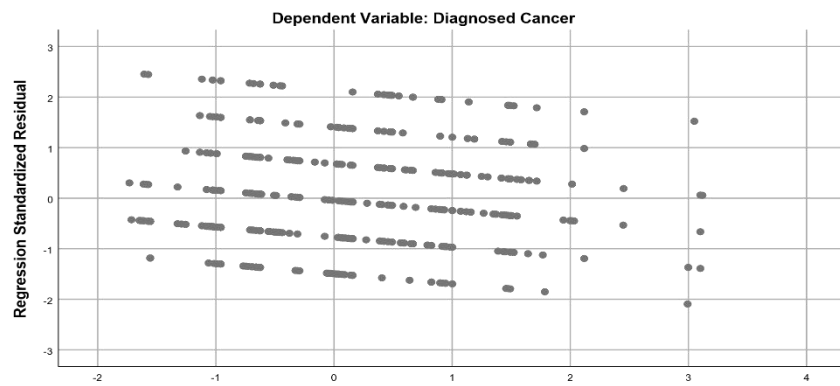
Normal P-P Plot of Regression Standardized Residual

Dependent Variable: Diagnosed Cancer



Observed Cum Prob Scatterplot

Chart 2: Regression Standardized Predicted Value



Regression Standardized Predicted Value

The analysis showed the diagnosed cancers are positively co-related with gender, occupation & idea about cause of cancer and negatively co-related with division, previously cancer history in relatives & associated disease.

4.10.4 Hypothesis testing

Null Hypothesis: The regression model is not significant.

Alternative Hypothesis: The regression model is significant.

Table 4.10.5: Determination of Hypothesis related to the regression model by ANNOVA test.

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	38.336	6	6.389	3.361	.003 ^b
	Residual	946.761	498	1.901		
	Total	985.097	504			

a. Dependent Variable: Diagnosed Cancer

b. Predictors: (Constant), Idea about Cause of Cancer, gender, Previous cancer history in relatives, Division, Occupation, Associated disease

The test is done to know which hypothesis is statistically important. It could measure by the significance value which is .003.

Result for multiple regression model: Here we see that the significance value is .003 which is less than .05. So the null hypothesis is rejected and alternative hypothesis is accepted.

As a result we can say that the regression model is significant.

4.11 Discussion

This is a comprehensive report on the burden of hematological malignancy in Bangladesh. In contrast to WHO estimate our data present a different picture. The most frequent hematological Malignancy is ALL (31.7%) which is one-third of total hematological malignancy cases where the AML is 20.6%, CML is 19.6%, NHL is 11.1%, HL is 10.7% & CLL is 6.3% [table-8]. According to the prediction of WHO, Non-Hodgkin Lymphoma is the commonest hematological malignancy & the rate of which is 1.9 per 100,000 persons. Chronologically 2nd most is leukemia which rate is 1.7 per 100,000 persons, 3rd is Hodgkin Lymphoma and 4th is multiple myeloma [11]. In Pakistan, NHL is the most prevalent type of HM [11]. In US, NHL is the commonest cancer among HM, which is 1.5 times that of all leukemia [12] . In other Asian countries including Japan, Korea and Singapore, NHL is the most frequent hematological malignancies [11] [13] . In our study, there are some unexpected discrepancies. Lacking of proper referral system might be the cause of these discrepancies. Lymphoma is a hematological disorder but a small number of patients might have been admitted to the medical oncology department.

Younger population is seemed to be afflicted by hematological malignancy in Bangladesh which is differ from western countries. At diagnosis, the median age was 48 years. But the real median age can be lower. Acute myeloid leukemia, Acute lymphoblastic leukemia, chronic myeloid leukemia, Hodgkin's lymphoma, Non Hodgkin lymphoma are found to occur in young

adults in which the median age is between 27 to 48 years(table-4). Another side, chronic lymphocytic leukemia & multiple myeloma are frequently occur in childhood age in Bangladesh & also in western counties. Gender is an important risk factor for hematological malignancy. In our study we see that men are more frequent than female in hematological malignancy & the ratio is 1.7:1. In other world-wide study, we also see that hematological malignancy is gender-skewed and men are more frequent than female. It may be due to gender discrimination as considering the socioeconomic status female cases were unreported in low income families. Men get more priority in seeking medical attention which could be another cause of unreported female cases. The higher prevalence of HM in males might be the result of increased exposure to environmental and occupational risk factors, smoking, alcohol consumption as well as different hormonal and genetic background of males and females [14]. Acute Myeloid Leukemia and Acute Lymphoblastic Leukemia are the most frequent HM in Bangladesh and these two constituted 52.3% (n = 505) of leukemia cases (Table 8).). The frequency of ALL is one & half times higher than that of AML in Bangladesh. The incidence of ALL is relatively common in Italy, US, Switzerland and Costa Rica [15] . ALL is predominantly a disease of Childhood in western countries. Chronic leukemia constituted 25.9% of all HM in Bangladesh. CLL is the third most common type (19.6%, n = 99), while CLL is the least frequent (6.3%, n = 32) HM (Table 8). The frequency of CLL is three times higher than CLL [Table-8].

Working environment & Residual environment are the commonest risk factor for cancer. In our study, we co-relate Hematological Malignancy with respondent occupation and we find that Student (44%) are most vulnerable to Hematological Malignancy [Table-3]. Most probably the causes behind this are radiation & radio wave discharge from their cell phone, laptop & other

electric device. We also find that most of the cancer patients are from Dhaka which about 35.2% and 2nd most from Chittagong is about 18.8%. We have already known that developed area has more radiation, benzene vapor, radio wave & other carcinogenic factor. So these probably the cause of increasing cancer patients in these two cities.

In this study, we also correlate Hematological Malignancy with Associated infection, previous cancer history in relatives and idea about cause of cancer. We find that associated infection and previous cancer history in relatives have no significant impact on Hematological Malignancy. But Idea about cause of cancer has significant impact on Hematological Malignancy. A large number of HM patient about 80.6% has no idea about the cause of cancer. As a result they were careless about cancer risk factor. So Government can take cancer awareness program in order to decrease cancer incidence in Bangladesh.

4.12 Summery

In this study, we wanted to find out risk factor of hematological cancer. Total 505 cases were analyze in this study where we find that Age, Gender, Occupation and proper idea about the cause Cancer have impact on hematological cancer incidence.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Findings & Contribution

This study is done on diagnosed hematological cancer patient to find out the risk factor and understand the patterns and distribution of Hematological cancer in Bangladesh. Here we find that Age, Gender, Occupation and proper idea about the cause Cancer as impact of hematological cancer. Its mean that these are the risk factor in Bangladesh for hematological cancer. Acute myeloid leukemia is the most frequent **HM** in Bangladesh. Male has higher prevalence of **HM** in relation to Female. Having no idea about the cause of cancer has more chances to occur **HM**. In case of occupation students have higher prevalence of **HM**. The age between 10-19 is more frequent for Hematological cancer. More investigations are needed to explain the epidemiology and biology of hematological malignancies in Bangladesh.

5.2 Limitation

Insufficient information in hospital record is the major limitation that we face during this research. As a result, we have to collect data directly from patients by face to face interview. Many patients feel hesitation to provide proper information. We handle it by repeated informing them about our research.

There are only few articles about hematological cancer of Bangladesh. Moreover, no government organized data on hematological cancer. These are another limitation in our research work.

5.3 Recommendations future works

In this research, we analyzed data to predict risk factor of hematological cancer. More investigation are needed to explain the epidemiology, genetics and biology of hematological cancer in Bangladesh.

References:

1. Hossain, M. S., Iqbal, M. S., Khan, M. A., Rabbani, M. G., Khatun, H., Munira, S., ... & Rahman, F. (2014). Diagnosed hematological malignancies in Bangladesh-a retrospective analysis of over 5000 cases from 10 specialized hospitals. *BMC cancer*, *14*(1), 438.
2. World Health Organization. Latest global cancer data: cancer burden rises to 18.1 million new cases and 9.6 million cancer deaths in 2018. Press Release 263. International Agency for Research on Cancer. Geneva, Switzerland, 12 September 2018.
3. WHO, "The number of new cancer cases is on the rise globally," World Health Organization, Geneva, Switzerland, 2018. <https://www.who.int/cancer/resources/keyfacts/en/>
4. Das, P., & Horton, R. (2013). Bangladesh: innovating for health. *The Lancet*, *382*(9906), 1681-1682.
5. Hussain, S. A., & Sullivan, R. (2013). Cancer control in Bangladesh. *Japanese journal of clinical oncology*, *43*(12), 1159-1169.
6. Hussain, S. M. A. (2013). Comprehensive update on cancer scenario of Bangladesh. *South Asian journal of cancer*, *2*(4), 279.
7. LEHTINEN, T., & LEHTINEN, M. (1998). Common and emerging infectious causes of hematological malignancies in the young. *Apmis*, *106*(1-6), 585-597.
8. Bhutani, M., Vora, A., Kumar, L., & Kochupillai, V. (2002). Lympho-hemopoietic malignancies in India. *Medical Oncology*, *19*(3), 141-150.
9. WHO, "Cancer prevention," World Health Organization, Geneva, 2017., 2017.
<https://www.who.int/cancer/prevention/en/>
10. A. M. Jahid, "Molecular cancer diagnosis in Bangladesh," The Daily Star, Dhaka., 2019.
<https://www.thedailystar.net/health/news/molecular-cancer-diagnosis-bangladesh-1712689>

11. GLOBOCAN, I. (2012). Estimated cancer incidence, mortality and prevalence worldwide in 2012. International Agency for Research on Cancer. *World Health Organization. Press Release*, (223).
12. Tarver, T. (2012). *Cancer facts & figures 2012*. American cancer society (ACS) Atlanta, GA: American Cancer Society, 2012. 66 p., pdf. Available from.
13. Park, H. J., Park, E. H., Jung, K. W., Kong, H. J., Won, Y. J., Lee, J. Y., ... & Park, S. (2012). Statistics of hematologic malignancies in Korea: incidence, prevalence and survival rates from 1999 to 2008. *The Korean journal of hematology*, 47(1), 28-38.
14. Lichtman, M. A. (2008). Battling the hematological malignancies: the 200 years' war. *The Oncologist*, 13(2), 126-138.
15. Redaelli, A., Laskin, B. L., Stephens, J. M., Botteman, M. F., & Pashos, C. L. (2005). A systematic literature review of the clinical and epidemiological burden of acute lymphoblastic leukaemia (ALL). *European journal of cancer care*, 14(1), 53-62
16. BRAC: Bangladesh Health System in Transition: Selected Articles (2008).
https://www.k4health.org/sites/default/files/bangladesh_health_system_in_transition.pdf.