



## Investigation of Prescription Data to Define the Disease Demography of Bangladesh

Md. Asiful Islam\*<sup>1</sup>, Christe Antora Chowdhury<sup>2</sup>

<sup>1</sup> Dept. of Computer Science, American International University-Bangladesh. Dhaka, Bangladesh.

<sup>2</sup> Dept. of Community Medicine, Popular Medical College and Hospital. Dhaka, Bangladesh.

### KEYWORDS

Epidemiology  
Demographic Analysis  
Public Health  
Prescription Data

### ABSTRACT

The prosperity of a country is highly dependent on its public health. Bangladesh is very keen to achieve a sustainable public health sector as a developing country. Bangladesh reached the goal of MDG (Millennium Development Goals) and pursued the goal of SDG (Sustainable Development Goals). Without sound public health, it is impossible to achieve SDG. So, Bangladesh is investing a good amount of money in the public health sector. As a part of this grand vision, this research has been conducted on the disease demography of Bangladesh. An interested pharmaceutical company collected prescription data from all administrative districts of Bangladesh, and weather and population data have been collected as a part of this work. An in-depth analysis of these data has been performed to discover insights into Bangladesh's disease demography. Finally, a set of recommendations has been presented to assist the public health sector of Bangladesh.

### ARTICLE HISTORY

Received 13 April 2022  
Received in revised form  
25 May 2022  
Accepted 28 May 2022  
Available online 29 May 2022

© 2022 The Authors. Published by Penteract Technology.

This is an open access article under the CC BY-NC 4.0 license (<https://creativecommons.org/licenses/by-nc/4.0/>).

## 1. INTRODUCTION

Although an inevitable event of human life, death is sought to be delayed and lingered for as long as possible. The standard parameter of well-being and assurance to the human race that death is far away is- Health. The World Health Organization claims that health is "a state of complete physical, mental and social well-being, not merely the absence of disease or infirmity" (WHO, 1946) [1]. Health has also been defined as the primary requirement to improve the quality of life [2].

However, in a developing country like Bangladesh, a huge amount of the population, especially those living in rural areas, have little or no access to healthcare facilities [3,33]. Also, notions of health are context-dependent; this is because human diseases exist with respect to people and people live in various social and cultural environments, classes, gender, ethnic groups, etc. So, a country like Bangladesh, which has been subjected to various natural disasters such as floods, cyclones, epidemics, etc., and has a highly polluted environment, is likely to have health crises that developed countries are not exposed to [3,37,41,42]. Thus, Bangladesh is in a unique position with limited accessibility and development of healthcare facilities but faces high health risks, significantly affecting the rural population.

Medicines have been used to treat diseases for thousands of years, earlier in the form of herbs and concoctions and later on as registered drugs. With the evolution of mankind, drugs have been formulated and sold commercially at pharmacies and hospitals based on prescriptions given by medical professionals.

All this information related to diseases, medicines, causes, symptoms, and all relevant areas have been recorded and collected as medical data. Health information technology has been extremely beneficial to health care experts, policymakers, developers, and consumers of healthcare. There has been a constant need to integrate scientific information into the system for progression [4]. But a developing country like Bangladesh has very few resources in health informatics. A research gap has been observed in the literature that there is no extensive analysis of the disease demography of Bangladesh. Most of the existing work focuses on a specific disease or administrative area. So, the need for a comprehensive disease demography analysis that includes all the administrative regions and diseases is very important. In the light of this requirement, this paper focuses on the following objective.

### General Objective

- To investigate disease demography in Bangladesh.

\*Corresponding author:

E-mail address: Md. Asiful Islam <asiful.islam@aiub.edu>.

2785-8901/ © 2022 The Authors. Published by Penteract Technology.

This is an open access article under the CC BY-NC 4.0 license (<https://creativecommons.org/licenses/by-nc/4.0/>).

### Specific Objective

- To identify the districts with the most and least cases of diseases.
- To identify and calculate the most occurred diseases and most prescribed drugs in Bangladesh over the last four years.
- To visualize the most diseases occurring in all the districts in Bangladesh.
- To investigate the trends of the most occurred diseases in the country.

This study aims to investigate how a correct understanding of disease dynamics can be predicted for Bangladesh, which is vital for health policy and planning. This study addresses the following question.

1. What are the districts with the most and least number of diseases?
2. What are the most occurred diseases and most prescribed drugs in Bangladesh over the four years?
3. What are the trends of the most occurred diseases in the country?

### 2. SIGNIFICANCE OF THIS STUDY

The perception of what disease, disability, or hazard is ever-changing. It is a kind of event or condition that hamper human health that mandates significant monitoring, recording, and analysis. To identify trends and patterns and achieve a greater understanding of the cause and effects of these diseases, extensive medical research has been carried out for years [5].

Subsequently, health research has high value to society. Because it is possible to gain some important information about disease trends and risk factors, treatment outcomes or public health interventions, functional abilities, patterns of care, and health care costs and use [30]. Hence, gathering all possible data related to human health, storing them efficiently, and analyzing them to generate patterns, trends, and models to improve human health and tackle epidemics or outbreaks that may occur in the future is a paramount interest [6,7]. With diseases occurring like mushrooms, drugs have been developed too. Simultaneously, mathematical models can be built to predict the future [8].

So, medical research has great potential for exploring the hidden patterns in the datasets of the medical domain with respect to diseases, symptoms, drugs, and their side effects. These patterns can be utilized for clinical diagnosis. From medical research, The WHO estimated that 12 million deaths occur every year due to heart diseases [9]. The WHO used this information to conclude that heart diseases were the primary reason behind the death of adults. Countries like India and USA were identified as ones where every 34 seconds, a person died due to heart disease.

Lyme disease, the most common vector-borne disease in the USA, was identified to have occurred over 25,000 times in 2009, and these figures were expected to increase [10]. Via extensive medical research, it was observed that in the Northeastern and Midwestern United States, Lyme disease was caused by the bacterium *Borrelia burgdorferi sensu stricto*

and transmitted to humans by *Ixodes scapularis*, primarily nymphs [11]. Furthermore, Lyme disease was dominant in May and July after contact with the vector during outdoor activities in tick habitat, which was considered closed-canopy deciduous and mixed forests [12, 43]. So, a relationship was found between humans affected and the density of *B. burgdorferi*-infected ticks.

Medical data research has also enabled the modeling of infectious diseases and enhanced the insights into the population, including individual-level disease transmission dynamics and risk factors, virulence, and Spatio-temporal patterns of disease transmission. From biostatistical methods to massive agent-based, biophysical, ordinary differential equation (ODE) to ecological-niche models have been used [13].

Thus, through the years, medical research has revealed relationships, reasons, solutions, and hence better tackling mechanisms in healthcare. All this information leads to a better understanding of diseases, thus increasing mortality.

### 3. CHALLENGES OF MEDICAL DATA RESEARCH

Medical data research is primarily dependent on the medical and other relevant data gathered. Research and analyses can only be drawn based on the data that has been collected. A common study design in medical research is to give patients some intervention and then observe what happens to them over time. For example, blood glucose concentrations may be measured several times after a glucose drink [14,44]. Through this observation, data is gathered.

On the other hand, surveillance of infectious diseases has been carried out in many countries. A disease surveillance system was designed in Japan from 1993 to 1997 and organized by the Ministry of Health and Welfare, involving around 2400 senile medical institutions (SNI) to detect epidemics in their early stages [15]. Almost 8% of the total pediatric hospitals and clinics in that area were a part of this and gathered all data needed and, in the long run, detected epidemics in advance.

Medical research is dependent on the medical data gathered. Most medical records, specifically those in developing countries like Bangladesh, are stored manually. Most of the data is stored in replicated data entries on charts, clinical trial forms, insurance claim forms, and such documents [16]. These entries are recorded from diverse sources and may be subjected to various errors.

It has also been widely argued that the researcher's background significantly impacts the study's course in qualitative medical data research. However, given that the data is valid and passed sampling stages to ensure accuracy, medical individuals can interpret and analyze it in their preferred domain [17]. So, preprocessing and curing data is an issue, but it can be interpreted in multiple ways once standardized.

Medical data research has determined that chronic non-communicable diseases (CNCDs) are reaching epidemic proportions worldwide. Diseases such as cardiovascular conditions, certain cancers, chronic respiratory conditions, and type-2 diabetes have all been classified under this category [18,38]. Also, non-communicable diseases (CNCDs) have

been regarded as a grand challenge [39,40]. These are fundamental scientific problems requiring considerable enhancement in the quality and deliverability of healthcare while reducing costs [48]. And in a developing country like Bangladesh, many people living under the poverty line do not have access to healthcare facilities. Thus, their records are eliminated from the system.

On the other hand, the available raw medical data are widely distributed, heterogeneous in nature, and voluminous [3]. They vary significantly in nature and need to be recorded accordingly. Also, different types of medical data have been gathered in various methods, making it difficult to unify them and reach a conclusive decision. Thus, these data must be collected efficiently and accurately. Without complete and accurate data, the research conducted will be questionable.

Data gathering is a challenge on its own, and in developing countries such as Bangladesh, this challenge aggravates. Due to limited resources, lack of infrastructure, poor management, standard recording procedures, inability to generate awareness, and a majority of the country's population not having access to proper healthcare facilities, medical data gathering has always been a challenge [21,31,32].

Despite being densely populated and subjected to various challenges such as poverty, overpopulation, corruption, and vulnerability to climate change, Bangladesh is undergoing a health transition and manifesting the double burden of disease attributable to the emergence of non-communicable diseases [19]. Bangladesh's policymakers started exploring personal benefits offered by the private sector due to the limited resources in the public sector [20,35].

Bangladesh has been exposed to a considerable amount of development, resulting in improved medical facilities [21]. With large reception of funding from international bodies and the introduction of IT in the medical sector, Bangladesh has room for offering better procedures for recording medical data and storing them efficiently.

#### 4. RATIONAL OF THIS STUDY

Thus, it has been possible to gather medical data, which is considered the local dataset for this research. The dataset comprises 7.6 million prescription entries, primarily focusing on Generic Name, Brand Name, and Disease Name. Concerning this dataset, accurate prediction of future observations and insight into the relationship between diseases and drugs for scientific purposes in quintessential [22].

This research investigates how a correct understanding of disease dynamics can be predicted for an under-developed country like Bangladesh, which is vital for health policy and planning.

#### 5. BACKGROUND

Demography is usually defined as studying human populations and their distribution, size, growth, density, and – statistics regarding birth, marriage, disease, and death [25]. Demographers have often suffered from 'identity' crises when locating themselves within a scientific discipline. By nature, demography is a multidisciplinary subject that emphasizes rigorous data analysis using specific methods accompanied by

the theory often associated with sociology, statistics, anthropology, economics, and public health. Health demography has become a discipline according to some scholars. Undoubtedly, demography's significant contribution to health policy comes from studying population dynamics in the determinants of mortality, fertility, and migration. Demographers have often analyzed family dynamics, kinship effects, and child development within these areas of study. Most of the time, demographers are more concerned with its implications for overall population size and structure and tend to ignore rare causes of death in the study of mortality. Sometimes it is not clear how demographers contribute to health policy debates and how healthcare policy in Bangladesh is defined. So, there is a clear need to understand and investigate the disease demography in Bangladesh through the collection and analysis of medical data.

Medical research is primarily dependent on medical data obtained. Nowadays, many concerns are directed toward hospital utility statistics, audits, resource allocation, vaccination updates, etc. [3]. Based on the information obtained from this research, patterns are identified such that diseases can be tackled better and their effects minimized or even eradicated. Bangladesh, being a country with minimal expenditure on healthcare facilities, medical research has not advanced much in gathering medical data. According to the latest Bangladesh National Health Accounts, US\$ 2.3 billion is spent on health or US\$ 16.20 per person per year, of which 64% comes through out-of-pocket payments. While according to WHO estimates, Bangladesh currently spends US\$ 26.60 per person on health per year [19]. This accounts that Bangladesh struggles to provide basic healthcare facilities to its citizens. Hence, minimum effort and funding have been allocated and even been possible for medical research; thus, the dire need to start now.

In general terms, medical data can be numerical statements about medical matters such as how many people die from a particular disease, mortality rate, etc., or in a descriptive manner such as symptoms observed, drugs prescribed, etc. [23]. Based on the data received, qualitative or quantitative analysis can be drawn by applying the principles of inductive reasoning and incorporating predetermined code types for data analysis and interpretation, generating taxonomy, theme, and theory [24].

Given recent developments and investments from foreign bodies, Bangladesh has made significant progress in developing its domestic pharmaceutical sector, introducing the National Drug Policy (NDP) in 1982. Domestic manufacturers now provide 75% of total drug sales and expand into developing an export market [19].

The local dataset takes into account prescriptions comprising such drugs. The Central Medical Store procures and distributes drugs to public sector hospitals and facilities. However, there is a chaotic market outside the public sector of some 64 000 licensed pharmacies and 70 000 unlicensed drug stores selling all medicines without requiring prescriptions. Records of these pharmacies are of significance too and need to be included in prospects [19].

These data obtained through various sources act as a basis for a great deal of medical research. The scope of medical research extends through specialties of medicines, research into diagnostic tests, prognostic factors, therapeutic and

prophylactic procedures, and covers public health and medical economies and clinical and epidemiological topics. Hence, it is essential to understand that medical data is not only relevant to specific patients or a singular disease. But health standards of entire nations or even continents, global health concerns, and similar issues are based on medical data obtained through research [14].

Early identification of the epidemic of infectious diseases is an outcome of successful medical research. This is an essential step toward proper intervention to control the disease and reduce its effects. Many contagious diseases are geographically and temporally limited by environmental variables such as temperature, humidity, rainfall, etc. Gathering detailed information on all these areas will better understand all contributing factors [26]. Therefore, collecting data in this domain is mandatory in Bangladesh since it is a country that has suffered from multiple epidemics and has vulnerable environmental conditions [19].

However, it has been argued that statistical evidence contributes to little or nothing in the progress of medicine because physicians and solutions are concerned with one patient at a time, and diagnoses vary depending on the patient's nature of illness, symptoms, and contraindications [23]. Despite this argumentation, statistical information on medical areas has been of crucial importance concerning evidence-based medicine, which describes the compilation of reliable and comprehensive information on medical care [27]. Thus, over the years, the aim has been to improve the reliability and credibility of medical research findings. An error in a published finding may influence all future aspects in that area; thus, a minor error stays dominant for years accounting for all future mistakes in that domain.

On the other hand, medical research in developed countries has excelled by miles compared to that in developing countries. Survival rates have been discovered via extracting knowledge from the data related to that disease. SEER (Surveillance Epidemiology and End Results) database, being a unique and reliable source of pre-processed data combining patient-level information regarding cancer site, tumor pathology, stage, and cause of death, has identified that the survival rate in breast cancer is 88% after five years of diagnosis and 80% after ten years of diagnosis [28]. These survival rates have shed new light on these areas; thus, data-driven scientific research has become an integral part of scientific areas such as medicine and biotechnology [29].

186,000 new cases of prostate cancer were diagnosed in America in 2008, with an estimated number of 28,000 deaths, making prostate cancer one of the most common malignancies in American men. Cancer screening trials in America have highlighted the overdiagnosis and overtreatment of the clinically insignificant tumor. As per data gathered from the European Randomized Study of Screening for Prostate Cancer, 48 cases of prostate cancer need to be treated to prevent a single death from prostate cancer within 10 to 12 years. This data has concluded that better tools and risk assessments for newly diagnosed prostate cancer patients are necessary [30].

Predictions models are also being constructed for the likelihood of infectious disease outbreaks and a better indication of how these diseases will spread and how to control them [45,46]. Methodologies such as risk factor

analysis, risk modeling, and dynamic modeling have been available due to medical research being conducted in these areas. Quantitative modeling of the epidemiology and control of human and animal diseases such as AIDS, influenza foot-and-mouth disease, and BSE has been identified as dimensions that need to be brought to the light, so the main drivers of disease risk from demography and behavior of land and climate change can be identified [34].

## 6. METHODS AND MATERIALS

A quantitative research method has been applied in this study. A descriptive and explorative study has been carried out. The total duration of the study was six months.

### 6.1 Source of Data

The primary data has been collected from a specific pharmaceutical company, the Bangladesh Bureau of Statistics, and the Bangladesh Meteorological Department.

### 6.2 Study Variable

**Table 1.** List of Variables

Independent Variable	Dependent Variable
Literacy rate	Disease profile
Weather	
Population	

### 6.3 Study Area

The study has been carried out on data collected from

- A specific pharmaceutical company based on prescriptions throughout Bangladesh.
- Literacy rate and population data from Bangladesh Bureau of Statistics.
- Weather statistics from Bangladesh Meteorological Department

### 6.4 Study Population

Almost 7.6 million data have been collected from the available sources.

### 6.5 Criteria for Sample Collection

Inclusion criteria:

- Only the data related to disease name, medicine prescribed, and location has been collected.
- Only the literacy rate and population-related data have been collected.
- Only the amount of rainfall, temperature, and location has been collected.

Exclusion criteria:

- No data other than that mentioned above has been collected for this research.

### 6.6 Data Analysis and Interpretation

The data has been analyzed in Python and MS-Excel. The analysis has been carried out in two steps. In Step 1: a disease profile was created. In step 2: the dependent variables were analyzed to investigate the effects of changes in independent variables on the dependent variable i.e., disease profile.

### 6.7 Outcomes

The expected outcome is to develop a disease profile for Bangladesh and identify the key reasons for the occurrence

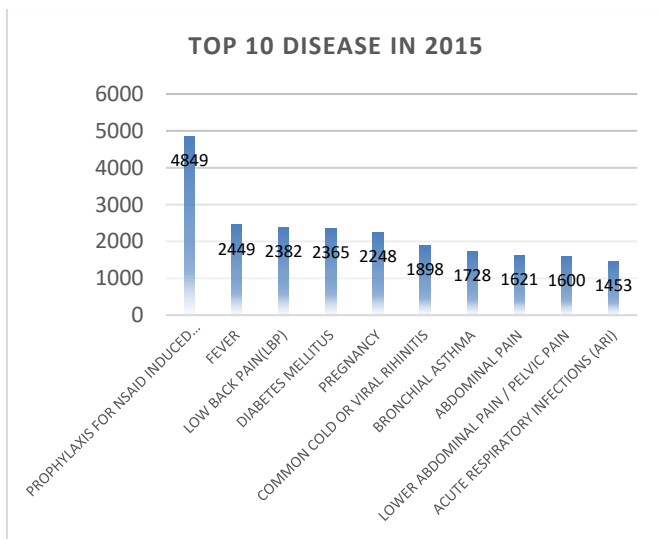
and distribution of disease in different locations in Bangladesh.

## 7. RESULT AND ANALYSIS

### 7.1 Disease Demography Analysis

#### Top 10 Disease in 2015, 2016, 2017, and 2018:

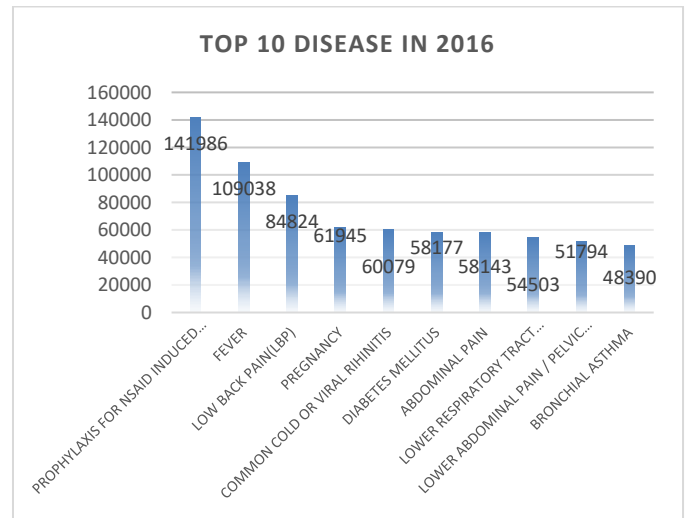
Figure 1 until Figure 4 represent the bar chart for the top ten diseases of Bangladesh in the years 2015, 2016, 2017, and 2018. The x-axis represents the disease name, and the y-axis represents the count of the total occurrence of a disease. Figure 1 to 4 shows that prophylaxis for NSAID-induced hyperacidity, fever, and lower back pain was the top three diseases in Bangladesh every year. Common cold or viral rhinitis, diabetes mellitus, and abdominal pain are top diseases in Bangladesh every year. Also, pregnancy is one of the significant cases found in the data.



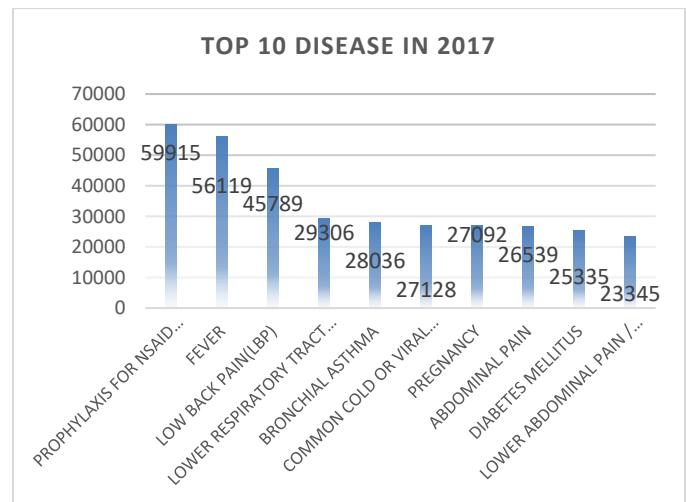
**Fig. 1.** Top 10 diseases in 2015

In 2015, prophylaxis for NSAID-induced hyperacidity was the top disease in Bangladesh. To be exact, 4849 occurrences of prophylaxis for NSAID-induced hyperacidity have been found in the data. Also, the second-most disease was fever, and fever occurrences were exactly 2449 in the data in the year 2015. And the 10<sup>th</sup> most occurred disease was acute watery diarrhea.

In 2016, prophylaxis for NSAID-induced hyperacidity was the top disease in Bangladesh. The exact count was 141986 occurrences. Also, the second-most occurred disease was fever, and fever occurrences were exactly 109038 in the data in the year 2016. And the 10<sup>th</sup> most occurred disease was bronchial asthma. The top three diseases of 2015 and 2016 were the same.



**Fig. 2.** Top 10 diseases in 2016



**Fig. 3.** Top 10 diseases in 2017

2017 shows an almost similar pattern to 2015 and 2016 for top diseases. The top disease for 2017 is prophylaxis for NSAID-induced hyperacidity. Fever and lower back pain were the second and third top diseases of this year.

Figure 4 also shows an almost similar pattern for 2018. The top disease was prophylaxis for NSAID-induced hyperacidity. Almost a similar number of people suffered from fever. Also, so many cases of the common cold or viral rhinitis, abdominal pain, and lower abdominal pain have been observed.

So, this analysis shows that every year most of the people of Bangladesh come to the doctor with the typical case of various pain, fever, common cold, and hyperacidity. So, preventive and preemptive measures for this common disease should be taken more. Also, enough medicine supply for common diseases should be ensured.

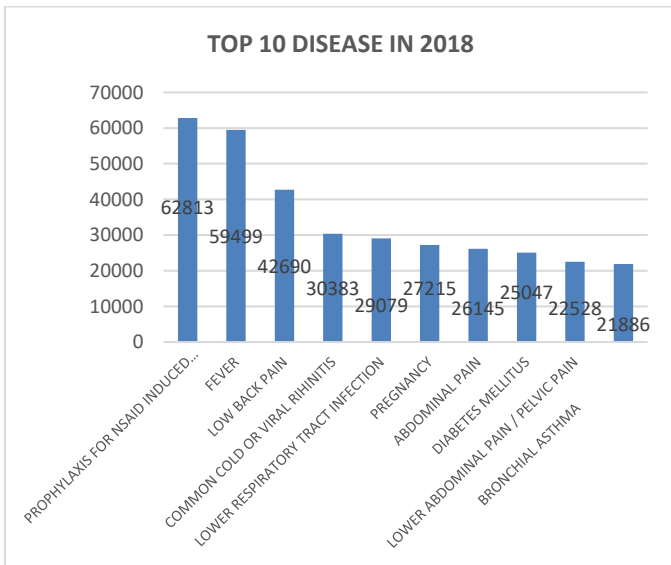


Fig. 4. Top 10 diseases in 2018

**Top 4 districts with the highest diseases:**

Figure 5 shows the top 4 districts with the highest disease, and Figure 6 shows the top 4 districts with the lowest disease.

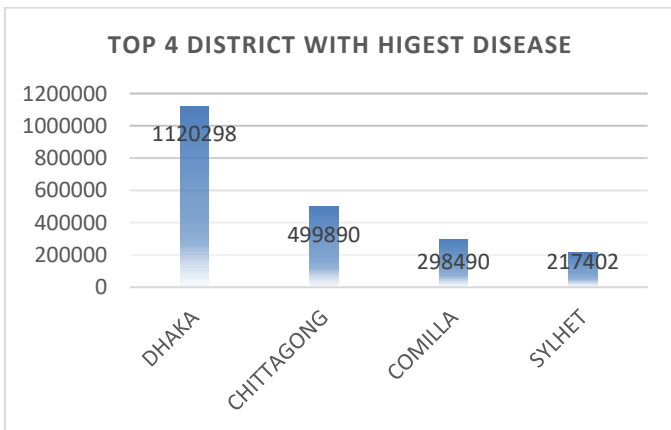


Fig. 5. Top 5 districts with the highest disease

From Figure 5, it can be seen that the highest diseases occur in the Dhaka district. Dhaka has the highest population density in the country. So, it makes sense that the highest number of diseases will occur in Dhaka. After Dhaka, Chittagong, Comilla, and Sylhet have the highest disease.

**Top 4 districts with the lowest diseases:**

The analysis of the Figure 6 shows that the lowest number of diseases occur in the Bandarban district. After Bandarban, Rangamati, Meherpur, Narail are three other district with the lowest diseases.

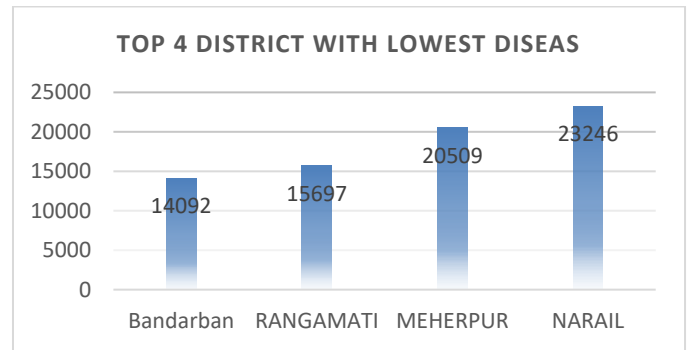


Fig. 6. Trendline for Fever

**Trend Line for top 3 diseases:**

The top three diseases that occur in Bangladesh are:

- Fever.
- Prophylaxis for NSAID-induced hyperacid.
- Low back pain(LBP)

Trendlines for these three diseases have been drawn based on the data from 2016 to 2018 and are represented in Figures 7, 8, and 9.

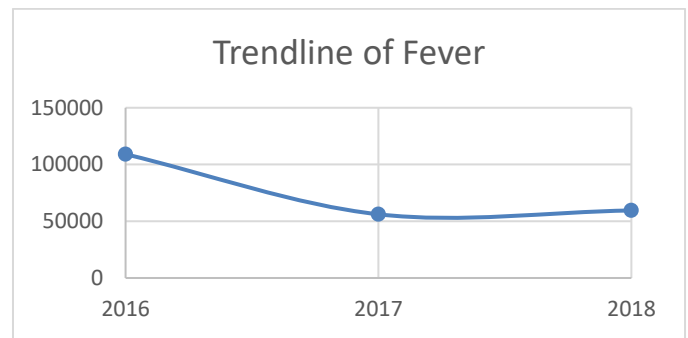


Fig. 7. Trendline for Fever

Trendline of fever shows that occurrence of fever reduced in 2017, and it slightly increased in 2018.

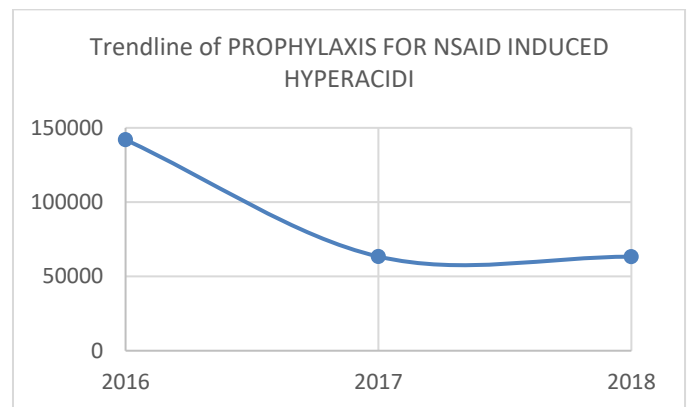


Fig. 8. Trendline for prophylaxis for NSAID-induced hyperacid

Prophylaxis for NSAID-induced hyperacid was also reduced in 2017. But in 2018, it was almost the same as in 2017.

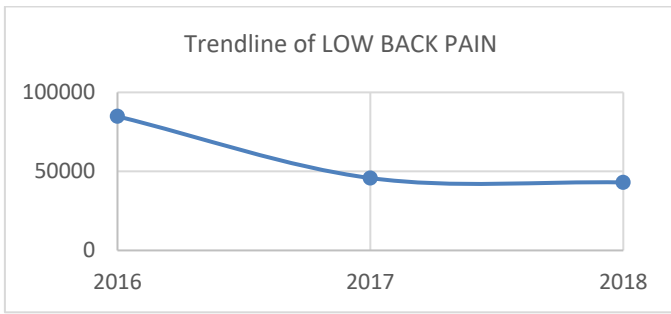


Fig. 9. Trendline for low back pain (LBP)

A dramatic reduction in lower back pain in 2017 has been observed. And the trendline shows that lower back pain (LBP) also reduced slightly in 2018.

**Top diseases in every district:**

The map in Figure 10 shows the top disease in each district. And color code represents different diseases in different regions. From the map, it can be seen that fever is the top disease in Dhaka, Chittagong, Barisal, and some other districts. Lower back pain is the top disease for Rangpur, Tangail, Jessor, Jhenaidah district.

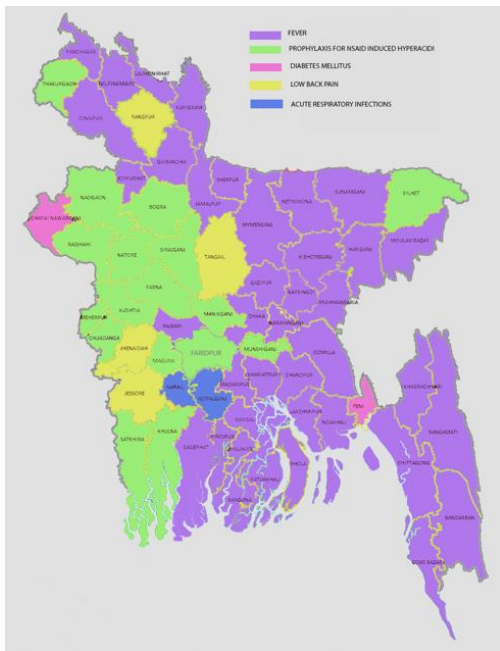


Fig. 10. Top Diseases in every district

**7.2 Drug Consumption Analysis**

**Top 10 used drugs in 2015, 2016, 2017, and 2018:**

Figure 11 to 14 represent four bar chart that shows the frequency of top drug uses in Bangladesh. The x-axis of the bar chart represents the drug name, and the y-axis represents the total count of the drugs prescribed. The generic name of the medicine has been used for this analysis.

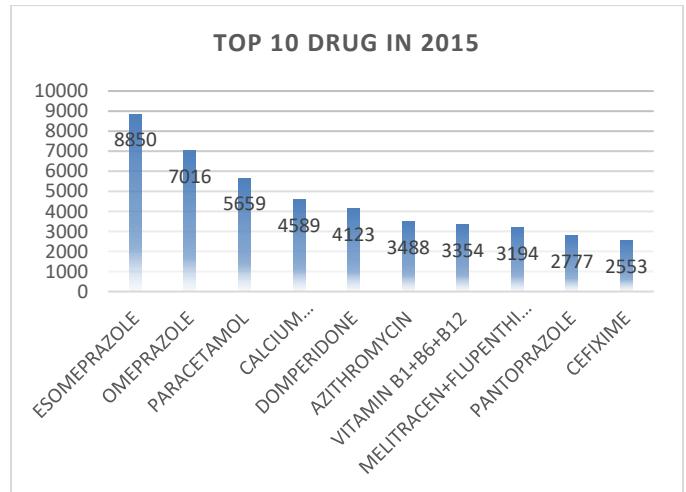


Fig. 11. Top 10 used drugs in 2015

In 2015, esomeprazole was the highest used medicine in the country. Esomeprazole is the medicine for stomach acid. Disease analysis in section 7.1 shows that the highest occurred disease is prophylaxis for NSAID-induced hyperacidity. So, it is logical that esomeprazole will be one of the highest used drugs in the country. Also, Omeprazole and paracetamol was the country's second and third highest used drug.

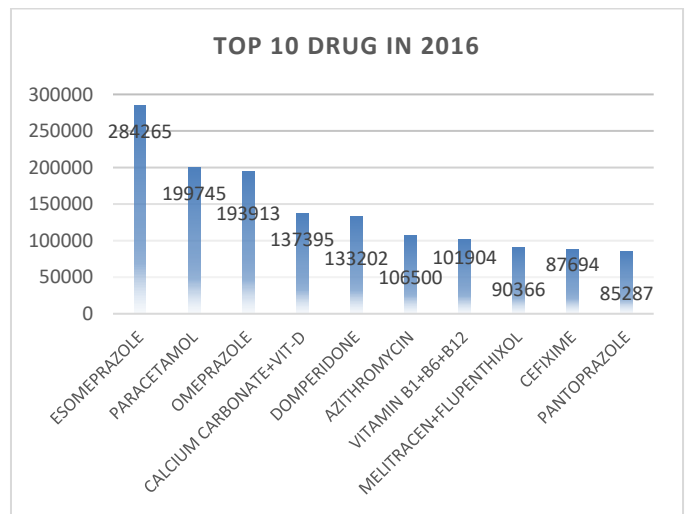


Fig. 12. Top 10 used drugs in 2016

2016 also shows almost the same pattern of drug usage. This year also, esomeprazole was the highest used, but paracetamol was the second-highest used drug in the country. Omeprazole, calcium carbonate + vitamin – d, domperidone, azithromycin are some top used drugs in 2016.

In 2017, esomeprazole was the highest used drug in the country. And the exact count of the use of esomeprazole was 140982. And the top ten used drugs pattern was almost similar to previous years.

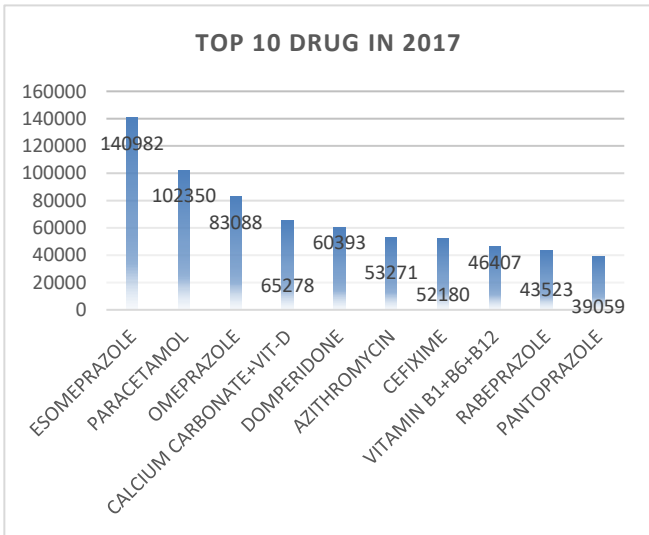


Fig. 13. Top 10 used drugs in 2017

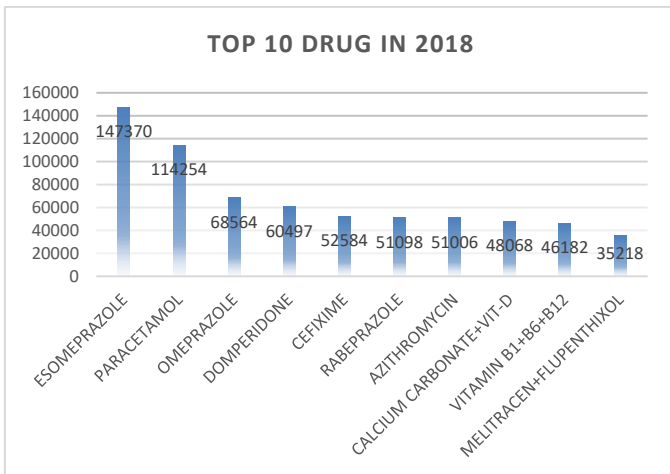


Fig. 14. Top 10 used drugs in 2018

From the analysis of Figure 14, it can be seen that esomeprazole was the highest used disease in the country in 2018. Also, paracetamol and omeprazole were the second and third highest used drugs. But the usage of rabeprazole was increased, and the use of cefixime was almost similar to the previous year.

From the analysis of these four years, It can be concluded that esomeprazole, paracetamol, omeprazole, and domperidone are some of the most frequently used drugs in Bangladesh. This analysis will be helpful for pharmaceutical companies to make various critical business and manufacturing decisions.

### 7.3 District Information Analysis

#### Data Collection:

For this analysis, two different types of data has been collected:

- i. Demographic data of districts.
- ii. Weather data of districts.

#### Demographic data collection method:

The following information about a district has been collected as a demographic data:

- Population.
- Literacy rate.

The Source of this demographic data is Banglapedia (<http://en.banglapedia.org>), the national encyclopedia of Bangladesh. A web crawler using python programming language has been developed to get the data from this website. The website was accessed on 23 April 2019 to get this data.

#### Weather data collection method:

The weather has a huge impact on human health [36]. Five years (2013 to 2018) of weather data has been collected from the Bangladesh meteorological department to find if there are any correlation between local weather and diseases.

Table 2. Partial view of district dataset

District	Disease Count	Population	Disease Weight (Disease Count/Population)	Population Density (Per Square km)	Literacy Rate (%)
DHAKA	1120298	8511228	0.1316	5685	67.15
CHITTAGONG	499890	6612140	0.0756	1252	55.55
COMILLA	298490	4595557	0.0650	1490	45.99
SYLHET	217402	2555566	0.0851	732	45.59
MYMENSING	182844	4489726	0.0407	1029	39.1
BARISAL	168598	2355967	0.0716	844	57
KISHOREGANJ	157978	2594954	0.0609	950	38.3
RAJSHAHI	153068	2286874	0.0669	950	47.54
NOAKHALI	142594	2640227	0.0540	715	52.32
BOGRA	140724	3013056	0.0467	1040	41.9

Table 3. Correlation matrix of the dataset

Correlation Matrix					
	Disease Count	Population	Disease Weight	Population Density	Literacy Rate
Disease Count	1				
Population	0.892219598	1			
Disease Weight	0.769033526	0.557688979	1		
Population Density	0.847369301	0.711944706	0.57386393	1	
Literacy Rate	0.400940333	0.268780548	0.44880773	0.33726825	1

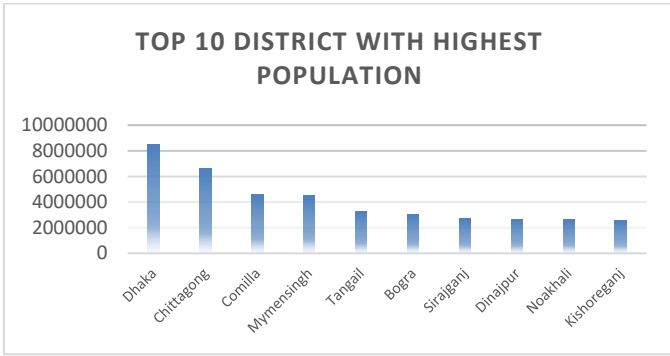


Fig. 15. Top 10 districts with the highest population

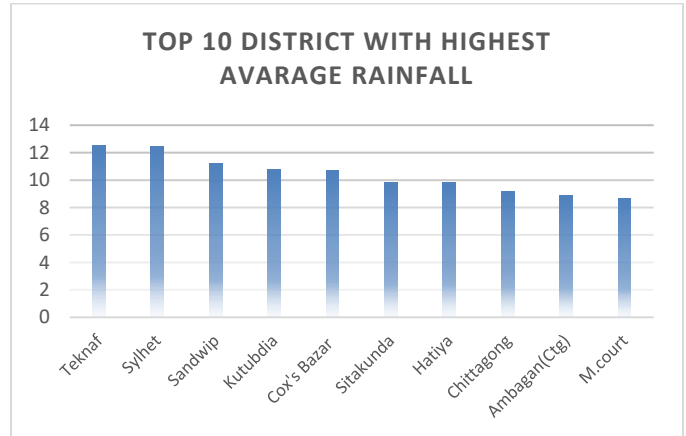


Fig. 19. Top 10 districts with the highest average rainfall

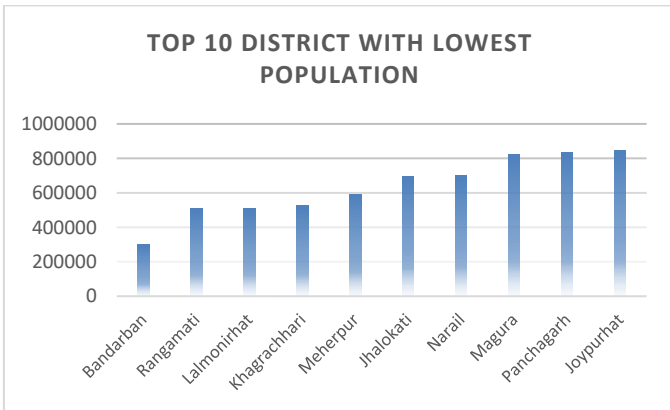


Fig. 16. Top 10 districts with the lowest population

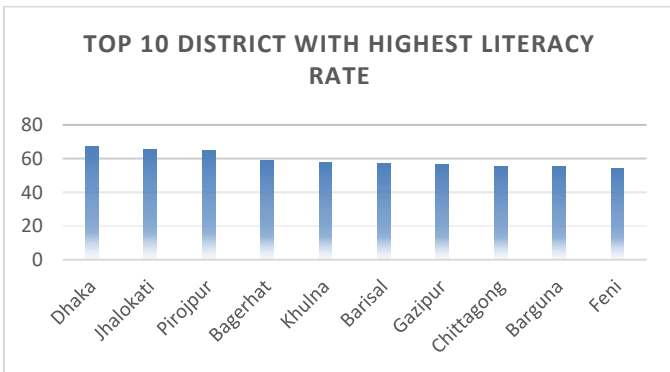


Fig. 17. Top 10 districts with the highest literacy rate

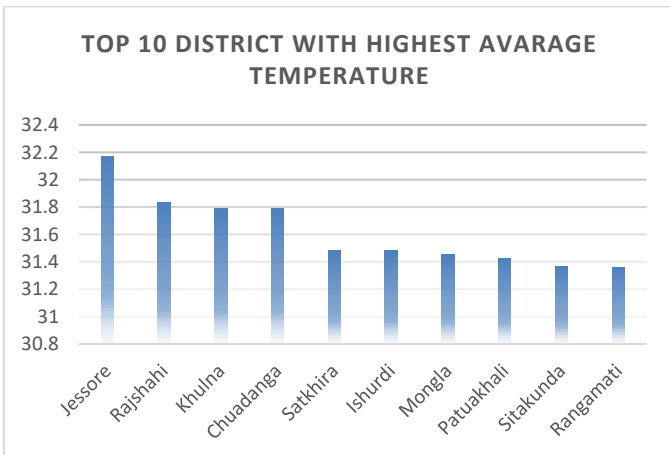


Fig. 18. Top 10 districts with the highest average temperature

### 8. DISCUSSION AND RECOMMENDATION

From the analysis of this paper, It can be seen that the most occurred diseases in Bangladesh are Prophylaxis for NSAID-induced Hyper Acidity, Fever, and Low Back Pain. Another important finding is that Diabetics is always present among the top 5 mostly occurred diseases in Bangladesh. From the medicine use analysis, It can be found that proton pump inhibitor types of medicines are frequently prescribed. Also, from the analysis of this paper, it is evident that antibiotics are frequently prescribed in Bangladesh. It is also alarming that a highly powerful antibiotic such as “Cefixime” is frequently prescribed. One interesting finding is that, apart from Dhaka and Chittagong, no other district with a high literacy rate features in the “mostly disease-prone list”. This finding shows the value of education. As though, the population rate has a direct effect on diseases. From weather analysis, it is evident that fever and rainfall are directly interrelated. The district “Chittagong” features in the list of “Highest Disease”, “Highest Temperature,” and “Highest Rainfall”.

Further detailed analysis is required to investigate the impact of weather, population, and literacy rate on disease. Based on the study, a comprehensive method should be developed to identify the disease trend in Bangladesh. Identifying disease trends will ensure the proper utilization of medical recourses in Bangladesh. It will also ensure that the impact of changes in population dynamics and climate can be minimized.

### 9. CONCLUSION AND FUTURE WORK

In this paper, a detailed analysis of the disease demography of Bangladesh has been presented. And this study showed some interesting correlations between illness and literacy rate. And some parameter of weather is also correlated with disease. This analysis will benefit the government for various kinds of decision-making in the health care sector. Also, medicine manufacturers, individual researchers, and various organizations related to the health care sector will get much important information for giving a better service. One crucial observation that has been discovered during this research is that the data collection mechanism for the health care sector is still not that much facilitated in Bangladesh. Even though a lot of data has been accessed from various sources for this research, it has been observed that many vital

parameters were missing. For example, this research would be more enriched the information regarding patients' age, gender, etc., could be found. So, there are many research opportunities in this field. In future research, we would like to investigate proper medical data collection mechanisms based on the available infrastructure in developing countries like Bangladesh. A good ecosystem of continuous medical data collection mechanisms will be helpful to keep track of the trend of the diseases. As a result, it will be possible to take adequate measures in advance to protect people from diseases. Overall, this research will be good guidance for the people who want to work on medical data analysis in Bangladesh.

## REFERENCES

- [1] Scully, Jackie Leach. "What is a disease? Disease, disability and their definitions." *EMBO reports* 5.7 (2004): 650-653.
- [2] Islam, Mohammad Shafiqul, and Mohammad Woli Ullah. "People's participation in health services: A study of Bangladesh's rural health complex." Bangladesh development research center (BDRC) (2009).
- [3] Soni, Jyoti, et al. "Predictive data mining for medical diagnosis: An overview of heart disease prediction." *International Journal of Computer Applications* 17.8 (2011): 43-48.
- [4] Chaudhry, Basit, et al. "Systematic review: impact of health information technology on quality, efficiency, and costs of medical care." *Annals of internal medicine* 144.10 (2006): 742-752.
- [5] Anne-Marie Yazbeck. "Health Systems and the Challenge of Communicable Disease Experiences from Europe and Latin America". *Central European Journal of Public Health* 17.1 (2009), p. 7
- [6] Gostin, Lawrence O., Laura A. Levit, and Sharyl J. Nass, eds. "Beyond the HIPAA privacy rule: enhancing privacy, improving health through research." (2009).
- [7] Sittig, Dean F. "Grand challenges in medical informatics?." *Journal of the American Medical Informatics Association* 1.5 (1994): 412.
- [8] Huppert, Amit, and Guy Katriel. "Mathematical modelling and prediction in infectious disease epidemiology." *Clinical microbiology and infection* 19.11 (2013): 999-1005.
- [9] Bell, Melanie L., et al. "Handling missing data in RCTs; a review of the top medical journals." *BMC medical research methodology* 14.1 (2014): 1-8.
- [10] Centers for Disease Control and Prevention (CDC). "Summary of notifiable diseases: United States, 2009." *MMWR. Morbidity and mortality weekly report* 58.53 (2011): 1-100.
- [11] Patnaude, Michael R., and Thomas N. Mather. Blacklegged Tick or Deer Tick, *Ixodes scapularis* Say (Arachnida: Acari: Ixodidae). University of Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, EDIS, 2003.
- [12] Pepin, Kim M., et al. "Geographic variation in the relationship between human Lyme disease incidence and density of infected host-seeking *Ixodes scapularis* nymphs in the Eastern United States." *The American journal of tropical medicine and hygiene* 86.6 (2012): 1062.
- [13] Corley, Courtney D., et al. "Disease prediction models and operational readiness." *PloS one* 9.3 (2014): e91989.
- [14] Matthews, JN1, et al. "Analysis of serial measurements in medical research." *British Medical Journal* 300.6719 (1990): 230-235.
- [15] Hashimoto, Shuji, et al. "Detection of epidemics in their early stage through infectious disease surveillance." *International journal of epidemiology* 29.5 (2000): 905-910.
- [16] Sung, Nancy S., et al. "Central challenges facing the national clinical research enterprise." *Jama* 289.10 (2003): 1278-1287.
- [17] Malterud, Kirsti. "Qualitative research: standards, challenges, and guidelines." *The lancet* 358.9280 (2001): 483-488.
- [18] Daar, Abdallah S., et al. "Grand challenges in chronic non-communicable diseases." *Nature* 450.7169 (2007): 494-496.
- [19] World Health Organization. Bangladesh health system review. Manila: WHO Regional Office for the Western Pacific, 2015.
- [20] N Kawnine, N and Guinness, L and Rahman, MA and Killingworth, JR and Wong, YH and Hussain. "private sector medical clinics and hospital survey". In: Retrieved from Ministry of Health and Welfare website: [http://heu.gov.bd/phocadownload/Researchn%20Paper%20No%2015\(1998\)](http://heu.gov.bd/phocadownload/Researchn%20Paper%20No%2015(1998)).
- [21] Khatun, Fatema, and Mst Khanam Sima. "Impact of ICT on health services in Bangladesh: a study on Hobiganj Adhunik Zila Sadar hospital." *Impact of ICT on Health Services in Bangladesh: A Study on Hobiganj Adhunik Zila Sadar Hospital (April 6, 2015)* (2015).
- [22] Fan, Jianqing, and Jinchi Lv. "Sure independence screening for ultrahigh dimensional feature space." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70.5 (2008): 849-911.
- [23] Armitage, Peter, Geoffrey Berry, and John Nigel Scott Matthews. *Statistical methods in medical research*. John Wiley & Sons, 2008.
- [24] Bradley, Elizabeth H., Leslie A. Curry, and Kelly J. Devers. "Qualitative data analysis for health services research: developing taxonomy, themes, and theory." *Health services research* 42.4 (2007): 1758-1772.
- [25] Leone, Tiziana. "How can demography inform health policy?." *Health Economics, Policy and Law* 5.1 (2010): 1-11.
- [26] WHO, Kuhn K., et al. "Using climate to predict infectious disease epidemics." *Communicable Diseases Surveillance and Response Protection of the Human Environment Roll Back Malaria*. Geneva, Switzerland: World Health Organization (2005).
- [27] Sackett, David L., et al. "Evidence based medicine: what it is and what it isn't." *Bmj* 312.7023 (1996): 71-72.
- [28] Smigal, C., R. Siegel, and A. Jemal. "Breast cancer facts & figures 2005-2006." *American cancer society*, Atlanta, Georgia (2007).
- [29] Bellaachia, Abdelghani, and Erhan Guven. "Predicting breast cancer survivability using data mining techniques." *Age* 58.13 (2006): 10-110.
- [30] Lowrance, William T., and Peter T. Scardino. "Predictive models for newly diagnosed prostate cancer patients." *Reviews in urology* 11.3 (2009): 117.
- [31] Hsu, Chih-Wei, Chih-Chung Chang, and Chih-Jen Lin. "A practical guide to support vector classification." (2003): 1396-1400.
- [32] Iqbal, Mohammed Masum, and Md Hossain. *Patients' Satisfaction with the Medical Services in Bangladesh: A Case Study on the City Hospital Limited, Dhaka*. 2015.
- [33] Ali, Kazi Julfikar, and A. Noman. "Determinants of demand for health care in Bangladesh: an econometric analysis." *3rd Asia-Pacific Business Research Conference*. 2013.
- [34] Woolhouse, Mark. "How to make predictions about future infectious disease risks." *Philosophical Transactions of the Royal Society B: Biological Sciences* 366.1573 (2011): 2045-2054.
- [35] Rahman, Redwanur. "Private sector healthcare in Bangladesh: Implications for social justice and the right to healthcare." *Global Public Health* 17.2 (2022): 285-296.
- [36] Cheng, Jian, et al. "Impacts of exposure to ambient temperature on burden of disease: a systematic review of epidemiological evidence." *International journal of biometeorology* 63.8 (2019): 1099-1115.
- [37] Biswas, Tuhin, et al. "Association between socioeconomic status and prevalence of non-communicable diseases risk factors and comorbidities in Bangladesh: findings from a nationwide cross-sectional survey." *BMJ open* 9.3 (2019): e025538.
- [38] Talukder, Ashis, and Md Hossain. "Prevalence of diabetes mellitus and its associated factors in Bangladesh: application of two-level logistic regression model." *Scientific Reports* 10.1 (2020): 1-7.
- [39] Riaz, Baizid Koorshid, et al. "Risk factors for non-communicable diseases in Bangladesh: findings of the population-based cross-sectional national survey 2018." *BMJ open* 10.11 (2020): e041334.
- [40] Al-Zubayer, Md Akib, et al. "Double and triple burden of non-communicable diseases and its determinants among adults in Bangladesh: Evidence from a recent demographic and health survey." *International Journal of Clinical Practice* 75.10 (2021): e14613.
- [41] Hennessey, Mathew, et al. "Intensification of fragility: Poultry production and distribution in Bangladesh and its implications for disease risk." *Preventive Veterinary Medicine* 191 (2021): 105367.

- [42] Gupta, Rajat Das, Promit Ananyo Chakraborty, and Md Belal Hossain. "Association of household wealth and education level with hypertension and diabetes among adults in Bangladesh: a propensity score-based analysis." *Tropical Medicine & International Health* 26.9 (2021): 1047-1056.
- [43] Nawrocki, C. C., and A. F. Hinckley. "Experiences with tick exposure, Lyme disease, and use of personal prevention methods for tick bites among members of the US population, 2013–2015." *Ticks and Tick-borne Diseases* 12.1 (2021): 101605.
- [44] Mathew, Thomas K., and Prasanna Tadi. "Blood glucose monitoring." *StatPearls* [Internet]. StatPearls Publishing, 2021.
- [45] Ding, Cheng, Xiaoxiao Liu, and Shigui Yang. "The value of infectious disease modeling and trend assessment: a public health perspective." *Expert Review of Anti-infective Therapy* 19.9 (2021): 1135-1145.
- [46] Randhawa, Nistara, et al. "Fine scale infectious disease modeling using satellite-derived data." *Scientific reports* 11.1 (2021): 1-11.