Determination of the Best Hospital (Hospital Advisor)

Team Control Number: **2018051**

Table of Contents

Summary

Anyone, at certain points in their lives, will have to visit the hospital for procedural checkups or treatments for accidents and diseases. In emergencies, patients are sent to the nearest hospital for immediate treatment within the "Golden Hour" of medicine. However, in the case of non-emergencies, individuals have the freedom to choose the hospital that best suits the patient and the area of treatment required. Certain hospitals have their specialities in specific areas of treatment and employ more specialized doctors in some departments than in others. Therefore, it is in the best interest of prospect patients to choose not just the "best" hospital but the best hospital for each individual's case. In this solution, we aim to quantify and model the various variables that affect the quality of a hospital in order to provide the user a ranked list of the best nearby hospitals.

To begin our research, we found the most prevalent disease categories and diseases in each category and collected the number of deaths, incidences and the average duration of stay for each disease. Then, we hope to provide information about which hospitals would best suit a patient by first developing the Patient Rating system (PR). The Patient Rating treats each patient's case, or visit to the hospital, as a dataset. In this dataset, factors such as the patient's disease, age, gender, and duration of stay in hospital are recorded. With the collection of a dataset for each patient, we can summarize and analyze a hospital's quality for specific diseases, ages, and genders.

Then, this compilation of data is used in our hospital rating system. Through the Hospital Rating (HR), we hope to create a balanced aggregate score that considers patient ratings, experience of doctors, doctor to patient ratio, patient reviews of hospitals, and more. Each of these various factors that contribute toward the overall hospital rating is given a unique weighting in order to account for the fact that some qualities of a hospital are more important to a patient than other factors. Eventually, the hospital rating will rank all of the hospitals being compared and the user will easily be able to pick the best hospital for them.

In order to process and simulate our model with the collected data, we utilized the Java programming language with the Processing 3 framework. Also, we plotted the resulting simulated data points in order to compare various factors with one another. Ultimately, this program allowed us to effectively compare a variety of data points from different data sets to determine which hospital would best suit a patient.

Glossary

Risk factor (RF): The estimated measure of how likely the patient will die due to the severity of the disease. See “Risk Factor (RF) Model” for mathematical definition.

Incidence: The estimated measure of the number of new cases of a disease that develops during a certain population and time period. In our model, incidence is measured in number of cases per year:

Incidence = Number of New Cases of a Disease

Error: A mistake caused by the hospital, due to mistreatment and/or mismanagement, which causes either a relapse or a development of a new illness/disease.

Duration of Stay (DOS): The time, in days, the patient stays in the hospital or attends for treatment until they are either in recovery (healthy discharge) or dead.

Inevitable Death: A death caused by a disease of a relatively high risk factor and low records of recovery.

Evitable Death: A disease that has a relatively low risk factor and a high chance of recovery.

Interpretation of Problem

During an non-emergency situation, a person has to compare different hospitals and eventually, choose the ‘best’ one. We are requested to help ease this process by developing mathematical models that will help evaluate the quality of the hospital.

The math model for the quality of the hospital will be composed of several variables, which represent the factors that help best judge the best hospital.

<u>Part 1: Mortality</u>

Using the mortality rate of the hospital to compare hospital quality is erroneous because one hospital could receive a significant number of patients with inevitable diseases than another hospital. This contrast would skew the data because as the mortality rate increases, the quality of the hospital decreases. With this type of model, hospitals in disease-stricken areas, for example, would receive very low ratings even if their treatments are proficient because their patients are inevitably bound to die. Therefore, the *amount of inevitable deaths* should not contribute towards the evaluation; rather, more focus must go to evitable deaths.

Also, We need to create a mathematical model that accounts for the severity of a disease and any procedural errors made by the hospital when looking at mortality.

Also, each treatment/death is made up of various elements that can contribute towards the hospital's quality. All of the different treatments/deaths need to be compiled in order to find a mathematical relationship between mortality and the hospital's quality. Hence, we need to create a dataset of various treatments of different diseases. Later, this dataset will help determine a rating system on which we evaluate the hospital.

<u>Part 2: Other Ideal Qualities of a Hospital</u>

Besides mortality, there are numerous other factors that could define the quality of a hospital. Out of the many different qualities expected in a hospital, we found that the following are the most important factors that affect the quality of the hospital:

- Doctor and staff experience
- Hospital capacity and staff numbers
- Attention given to patient
- Past patient reviews

Now, we need to create individual models or datasets for these factors in order to find a way to record/calculate them. For example, each doctor in a hospital can be treated with a dataset that holds information about what they specialize in and how many years they have been working as a doctor. After that, we are required to integrate these factors into our main model, which will eventually determine the best hospital for the patient.

Assumptions

1. The patient's budget does not restrict the location of treatment.
    - ➢ Even with health insurance, medical care and treatment is often expensive, especially in the United States. However, we assume the patient wishes to receive the most safe and reliable treatment possible, regardless of its cost.

2. When requested, hospitals are able to provide data that pertains to patients' information of hospitalization.

- ➢ This data includes: i) the duration of stay (D.O.S.) of a patient with a certain disease. ii) the age and gender of patients. iii) the primary and final diagnosis of patients. These factors are used in our model to determine the quality of hospitals, but the general public do not have open access to these confidential data. To demonstrate our model, we gathered as much data as possible from different sources and time periods (this is further elaborated upon in the Strengths and Weaknesses section). We assume that hospitals will trust and provide companies that wish to create a hospital ranking program with sufficient data.

3. All of the hospitals provide basic necessities required for treatment. These include food, water and clean, modern and maintained medical equipment and facilities.
   - ➢ Every patient in hospital needs to receive water and food. Also, each patient needs to be treated and/or operated upon with the cleanest and well maintain medical equipment. Hence, we assume each hospital is able to provide their patients with these basic necessities.

Models and Justification

Part 1: Patient Rating and Mortality Model

Mortality needs to be evaluated using smaller factors such as severity of disease, duration of stay, and any procedural errors made by the hospital. This is done to judge whether a death was evitable or inevitable, and to evaluate the hospital's treatment at the same time. In order to best judge a hospital based on these mortality factors, each patient's treatment is treated as a dataset and assigned a 'Patient Rating' value (PR). Each new treatment (excluding relapse or hospital error) will be assigned a new PR value based on calculations which vary depending on the treatment by the hospital. Below is an example of what a hospital would look like holding its patient datasets (this hospital has 2 patients):

Hospital X	**Age**	**Gender**	**Disease**	**Risk Factor (RF)**	**Duration of Stay (DOS)**	**Result**	**Patient Rating**
Patient	46	F	Prostate	56.86%	4 days	Recovery	0.74

#1			Cancer				
Patient #2	69	M	Liver Cancer	12.45%	7 days	Death	0.12
Average Patient Rating:							0.43

Disregarding RF and DOS as of now, this example illustrates that it is advantageous to hold data this way, as we can compare two categories such as disease and PR to see which disease the hospital is most successful in treating. This will be further elaborated upon in “Hospital Rating” and “Data Comparison for Better User Compatibility.” For now, the question is how one calculates the PR, which aims to measure the quality of treatment a patient receives at a hospital by only observing mortality factors.

The PR was intended to fall under a certain interval. For this we set a PR value that starts at 0.5 for each patient and it was to be changed based on the mortality factors previously defined. The changes made to that initial 0.5 were modeled in such a way that most of the PR values fell in between 1 and 0, so that 0.5 is the standard point given to a patient that gets standard treatment (exceptions will be talked about in “Risk Factor (RF)”). This change is denoted as ΔPR (PR = 0.5 + ΔPR) and is affected by three factors: Risk Factor (RF), Duration of Stay (DOS), and whether the patient recovered or passed away.

The following equation illustrates that ΔPR is the sum of its two factors.

$$\Delta PR = \Delta PR_{RF} + \Delta PR_{DOS}$$

I. where PR_{RF} is the change in the PR value affected by the magnitude of the patient’s risk factor and whether the patient recovered or passed away.
II. where PR_{DOS} is the change in the PR value affected by comparing the patient’s duration of stay to the average duration of stay of the disease and whether the patient recovered or passed away.

If an error by the hospital is determined, a slightly different application of Equation 1 is used. See “Error (E) model”.

Risk Factor (RF) Model

When approaching this problem, our team first dealt with the problem of how to differentiate between evitable and inevitable deaths. Our model used very specific data and we wanted to use that to our advantage. We knew that every patient held information about what disease they were carrying, and whether they recovered or died. So, we decided to give each disease an RF value so that evitable and inevitable deaths can be seen as a spectrum. The higher risk a disease was, the more inevitable a death was from it. The lower a risk a disease was, the more evitable a death was from it. To estimate the severity of a disease we came up with the idea of creating a ratio between the incidence of a disease in a given year and the number deaths due to that disease in the same year.

$$\text{RF} = \frac{\textit{Number of Worldwide Deaths from a Disease}}{\textit{Number of Worldwide Cases of the Disease}}$$

Then, we classified diseases into different categories and collected data on their incidence and death values for a given year. This enabled us to calculate the risk factor for our database of diseases. Below is the data we obtained:

Collected Data

Disease Category	**Disease**	**# of Cases (Entire World)**	**# of Deaths (Entire World)**	**Risk Factor (%)**
CVD	Coronary Heart(Ischaemic)	79800000.00	8756000.00	10.97
	Stroke	9000000.00	6241000.00	69.34
	Congenital Heart	5700000.00	302000.00	5.30
Cancer	Lung Cancer	1448000.00	1415081.00	97.73
	Liver Cancer	6328000.00	788000.00	12.45
	Colorectal Cancer	1080000.00	744000.00	68.89
	Stomach Cancer	933000.00	754000.00	80.81
	Prostate Cancer	605000.00	344000.00	56.86
	Cervix Cancer	489000.00	279000.00	57.06
	Breast Cancer	1100000.00	571000.00	51.91
	Leukemia	375000	290000.00	77.33
	Lymphomas	479000.00	210000.00	43.84
Diabetes	Diabetes	422000000.00	1585530.00	0.38
Neurologic Disorders	Alzheimer + Other Dementias	35600000.00	1541880.00	4.33
	Parkinson	10000000.00	128399.00	1.28
	Epilepsy	50000000.00	152802.00	0.31
	Multiple sclerosis	2500000.00	22984.00	0.92
	Migraine	Not Needed	0.00	0.00

	Non-Migraine Headache	Not Needed	0.00	0.00
Mental Disorders	Depression	Not Needed	0.00	0.00
	Bipolar Disorder	Not Needed	0.00	0.00
	Anxiety Disorder	Not Needed	0.00	0.00
	Somatic Disorder	Not Needed	0.00	0.00
Substance Abuse	Alcohol Use	5000000.00	128910.00	2.58
	Drug Use	5000000.00	167750.00	3.36
Respiratory Diseases	Asthma	234900000.00	383176.00	0.16
Infectious and Parasitic Diseases	Tuberculosis	7800000.00	1373000.00	17.60
	HIV/AIDS	36700000	1060000.00	2.89
	Hepatitis	71000000.00	145074.00	0.20

Now that we obtained the data for various risk factors, we have to determine the relationship between the RF and the PR values. There are two main cases. The first is that the patient dies, and the other is if the patient recovers. If the patient dies the PR is going to decrease, if the patient lives the PR is going to increase. If the patient dies due to a disease with high risk factor (more inevitable), the PR is decreased by a small amount. If the patient dies due to a disease with low risk factor (less inevitable), the PR is decreased by large amount. If the patient recovers from a low risk disease (normal treatment), the PR is increased by a large amount. If the patient recovers from a high risk disease (well treated), the PR is increased by a large amount. These relationships are put into the table below:

Relationships between Risk Factors, Mortality and Patient Ratings (PR)

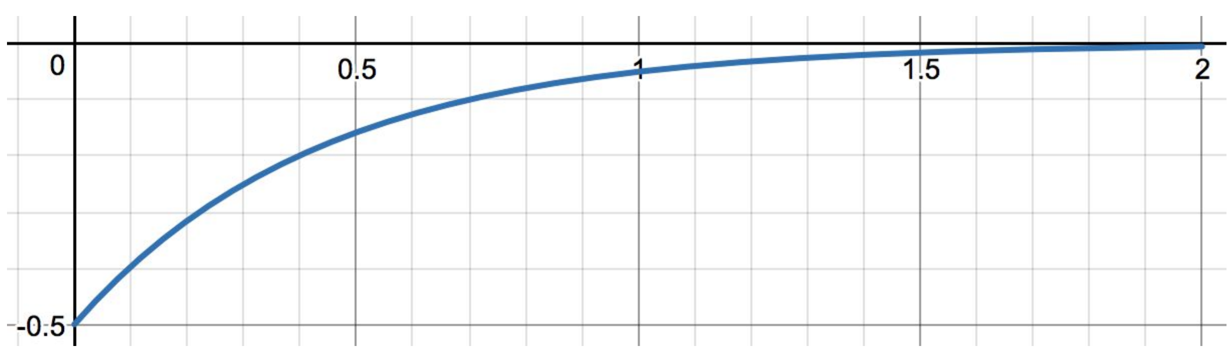
Risk Factor (RF)	**Mortality**	**Patient Rating (PR)**
High	Deceased	Down by Small
High	Recovered	Up by Big
Low	Deceased	Down by Big
Low	Recovered	Up by Small

These relationships considered, these equations were crafted. The choice of using exponential functions came from the fact that as the risk factor gets higher, the patient would be harder to treat.

$$\Delta PR_{RF} = 1.5^{RF/100} - 1 \text{, if the patient lives.}$$
$$\Delta PR_{RF} = 0.1^{RF/100} * (-0.5) \text{, if the patient dies.}$$

Patient Lives

Patient Dies



Note: These functions implicate that the PR value can be above 1. We allow this to happen but we do not allow for a PR value can to be below 0. If we find that a PR is less than 0, the PR for that patient is set to 0.

Duration of Stay (DOS) Model

Duration of Stay was created with the mindset that it would be a less impactful factor than Risk Factor.

This model should prioritize the success of treatment, so there are different functions for the duration of stay depending on whether the patient lived or died. Then, each function was modeled so that if a patient dies, the longer the hospital is able to keep the patient alive led to an increase in patient rating, as their death was delayed. If the patient lives, the less time the hospital takes to treat the patient leads to an increase in patient rating, as the patient was treated more efficiently. To be able to rate duration of stay, we collected data on the average duration of stay for certain diseases. When calculating the PR, the patient's duration of stay will be compared against the average DOS for the disease they are diagnosed with. The next two tables show the relationship between DOS vs. PR, and the Average DOS for certain diseases.

Relationships Between Treatment Result and Average Duration of Stay in Hospital

	Above Average D.O.S	**Average D.O.S**	**Below Average D.O.S**
Recovery	ΔPR: Negative Value	ΔPR: 0	ΔPR: Positive Value
Deceased	ΔPR: Positive Value	ΔPR: 0	ΔPR: Negative Value

Collected Data

Category	**Disease**	**Average D.O.S (Days)**
	Coronary Heart(Ischaemic)	4.10
	Stroke	5.10
CVD	Congenital Heart	5.60
	Lung Cancer	7.50
	Liver Cancer	6.90
	Colorectal Cancer	6.80
	Stomach Cancer	10.50
	Prostate Cancer	2.80
	Cervix Cancer	3.70
	Breast Cancer	2.60
	Leukemia	14.20
Cancer	Lymphomas	9.70
Diabetes	Diabetes	2.70
	Alzheimer + Other Dementias	4.90
Neurologic Disorder	Parkinson	4.60

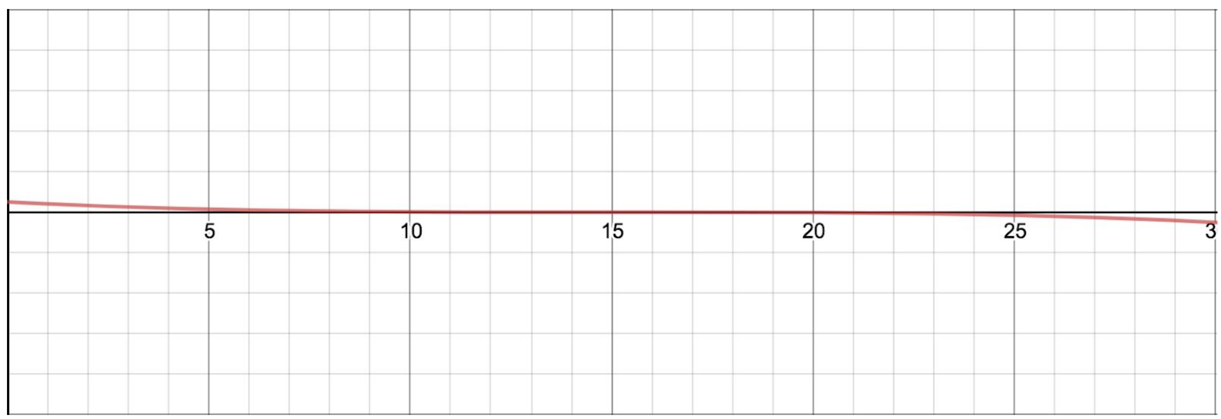
	Epilepsy	3.60
	Multiple sclerosis	4.90
	Migraine	2.80
	Non-Migraine Headache	2.80
Mental Disorders	Depression	2.80
	Bipolar Disorder	2.80
	Anxiety Disorder	5.20
	Somatic Disorder	5.20
Substance Abuse	Alcohol Use	6.40
	Drug Use	3.00
Respiratory Diseases	Asthma	3.50
Infectious and Parasitic Diseases	Tuberculosis	15.00
	HIV/AIDS	9.30
	Hepatitis	5.00

Keeping these relationships in mind, the equation below was created for how much a patient's DOS changes the PR.

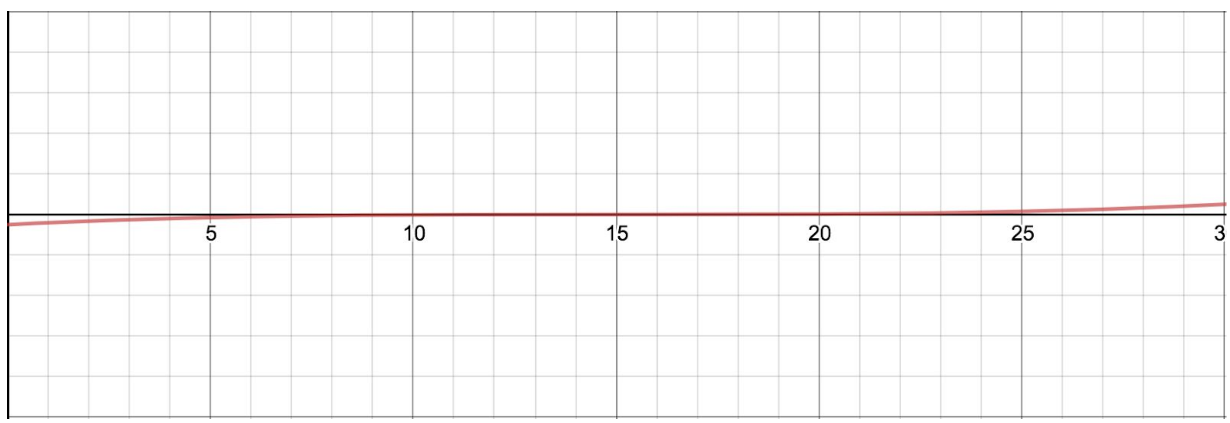
$$\Delta PR_{DOS} = \frac{1}{4(DOS_{avg})^3}(DOS - DOS_{avg})^3 * A$$

Where A = -1 if the patient lives, and A = 1 if the patient dies, DOS = patient's duration of stay in the hospital for treatment, DOS_{avg} = average duration of stay for patients'. We chose this relationship we felt that it would make sense if PR would change at higher rates as DOS got further from DOS_{avg}. Also, since we wanted DOS to have a lesser effect than RF, we set the made it so that most of the time $-0.25 < \Delta PR < 0.25$. This explains the coefficient of ¼. Below are the two graphs for a disease with an average DOS of 15 days, depending on the patient's case.

Patient Survives



Patient Dies



<u>Error (E) Model</u>

If it is determined that an error by the doctor or hospital occurs, the previously determined patient rating is reset to a constant, E. This error includes avoidable relapse and avoidable mistakes in treatments and surgeries. If the patient returns to the hospital to be treated for the error, Equation 1 is used for a new ΔPR. This new ΔPR is added to the constant E instead of the previously used value of 0.5.

If a patient lives even with the error, the error constant E=0.25. If a patient dies due to the error, E=0.01. When considering the readmission to the hospital, the E is always 0.25 because the ΔPR will likely drop close to 0 if the patient does not survive.

Part 2: Hospital Rating (Other factors)

Mortality cannot be the only factor that determines the quality of a hospital. Other factors that will contribute to the hospital's quality rating will be labelled as the Hospital Rating (HR). The 'Hospital Rating' is a value that is assigned to each hospital being compared to each other. Also, this rating system provides a method to help rank the hospitals from "best" to "worst." The HR value is comprised of various factors such as the Doctor to Patient Ratio (DP), Hospital Capacity Ratio (C), Doctor Medical Experience (DE), User Review (UR), and Patient Rating (PR). Similar to the Patient Rating system, each of these factors are weighted differently (as indicated in table below). Also, the HR value will change every year based on the data collected from the hospitals.

Factor	**Weight (%)**
Doctor to Patient Ratio (DP)	20
Capacity Ratio (C)	10
Doctor Experience (DE)	20
User Review (UR)	20
Patient Rating (PR)	30

The following equation illustrates that the 'Hospital Rating' is a weighted sum of its factors. Also, the range of the HR is $0 < HR < 1$.

$$HR = 0.1*C + 0.2*DP + 0.2*DE + 0.2*UR + 0.3*PR$$

Doctor to Patient Ratio (DP) Model

An important factor while evaluating the hospital's quality is the ratio between the number of doctors to the number of patients, as defined by the equation below.

$$DP = \frac{Number\ of\ Doctors}{Number\ of\ Patients}$$

This model is particularly important because it helps to show how much attention the hospital can give to its patients. If the DP ratio is low, it signifies that there are more patients per doctor; hence, each doctor will pay relatively less attention towards each patient. As a result, this will contribute a decrease to the hospital's quality, putting the particular hospital in a lower ranking.

Capacity Ratio (C) Model

The Capacity Ratio (C) is another vital factor in evaluating hospital quality. It is defined as the ratio between the number of patients to the maximum capacity of the hospital.

$$C = 1 - \frac{Number\ of\ Patients}{Max.\ Capacity\ of\ Hospital}$$

This model helps to judge a hospital's quality as it illustrates how 'full' the hospital usually is on a yearly basis. Similar to the DP ratio, this model also contributes towards determining whether or not a hospital is able to give ample attention to its patients. If a hospital is busier during the year, its C value is higher than other hospitals; hence, this will contribute a decrease to the HR value, putting the particular in a lower ranking.

Factors Defined by Data (DE, UR, PR)

The remaining three factors (Doctor Experience, User Review and Patient Ratings) are also datasets comprised of various data points.

For Doctor Experience (DE), a data set will be collected from the hospital regarding the work experience of each employed doctor or physician from each department. These work experience data points will be categorized by a doctor's or physician's department.

For User Reviews (UR), a data set will be generated based on the amount of virtual 'likes' a hospital receives in a given year. We need the User Review factor to help evaluate the hospital's quality based on patients' past experiences with the hospital.

For Patient Rating (PR), a data set will be generated from the aforementioned model (See Part 1: Patient Rating and Mortality Model). The Patient Rating (PR) helps illustrates how well the hospital prevents evitable deaths and treats diseases.

Later, these three distinct datasets, in addition to the Doctor to Patient (DP) and Hospital Capacity (C) ratio, will be computed into a single Hospital Rating (HR) through our Mapping Model.

Mapping Model

The factors defined by data (DE, UR, and PR) all have to be integrated into the Hospital Rating system. This can be done using the mapping function seen below:

$$\text{Value} = \frac{1}{H-L}(x-L)$$

I. where H is the highest possible value in the dataset
II. where L is the lowest possible value in the dataset
III. where x is any specific data point

This Mapping Model allows us to convert all data point values in the datasets for DE, UR, and PR into values that fit in the range of HR ($0 < HR < 1$). A new dataset will be created for each of the three factors (DE, UR, and DP) with values that are between 0 and 1. Since the new values are in between 0 and 1, and these are the values that end up being plugged into the HR equation to calculate HR. The purpose of this function is to create a ranking of sorts. Since UR, DE and DP are relative (there is no optimal value; the more you have the better it is) it is better to have all of those data points normalized to a value between 0 and 1 to be added to the weighted average to calculate HR.

Data Comparison for Best User Compatibility

Since our model treats each patient entry as a new dataset, we have the ability to compare multiple factors in different hospitals, allowing us to easily compare trends. Each patient entry holds information regarding the patient's Age, Gender, Duration of Stay, Primary Diagnosis (RF and readmission due to error), and Final Diagnosis (if the patient has recovered or passed away). The RF and DOS are used to calculate the PR. The two other factors, age and gender, are recorded to provide a basis of comparison for the user's compatibility with a certain hospital.

For example, imagine that a 30 year old is comparing the two best hospitals near his residence. If he finds that the HR are remotely the same or close to each other, the user can choose to compare Patient Ratings with other data tied to other hospital patients. The hospital that has the best PRs for their age group (30-39) is going to be more compatible for them. The same logic applies for the remaining factors that are listed in the previous paragraph.

This calls for the need of a new value: the **User Compatibility Rating (UCR)**. The next sections will display the calculation of this value for certain comparisons made with patient data. An overall UCR value can be calculated using the following equation (assuming we are only comparing the data points discussed in the following subsections):

$$UCR = CM * (\Delta UCR_{A} + \Delta UCR_{DC} + \Delta UCR_{DE})$$

(This equation will be talked about in the following subsections)

Note: Doctor data also needs to be added to each hospital in a similar way to the patient data as seen in the "Models and Justification" section of the paper. This is so that we can compare Doctor Experience to overall HRs.

Synthesis of Data via Simulation

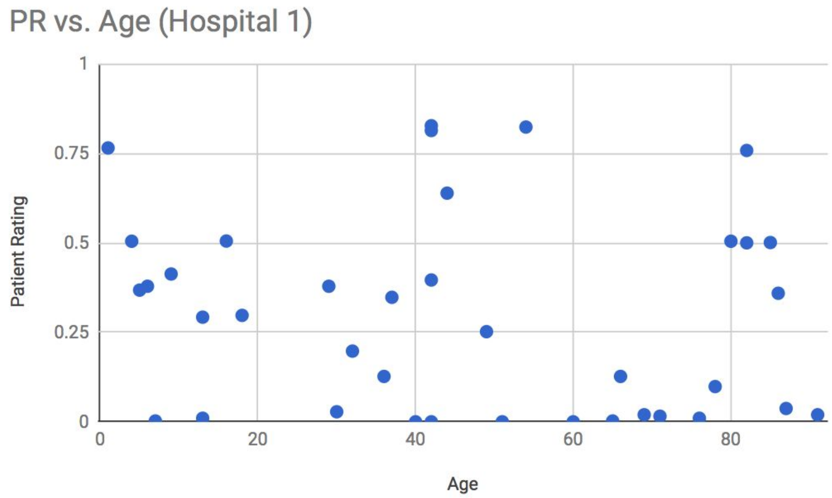
For the graphs and data you are about to see, we ran computer simulations to synthesize them. We followed our model exactly and gave patient datasets all the pieces of data depending on our collected information about diseases, RFs, and DOSs, and calculated Patient Ratings for each. We then created doctor datasets and gave them information about their specialties, and years of experience. Next, both patients and doctor datasets were compiled into hospitals as lists of datasets, and that was how HRs were calculated.

The coding process was done in Processing 3, which is a framework for Java that has easy to use library functions. The code consists of 5 classes. As mentioned before, each patient's treatment or death is a dataset. Similarly, each doctor is a dataset for each hospital. This "hierarchy" required multiple classes to work with each other. The 5 class hierarchy includes the following: Hospital Advisor (main class), Doctor, Hospital, Patient, Writer. The Doctor and Patient classes hold data pertaining to the doctors and patients of each hospital. These classes have functions to pass on that data to other classes, in which the visible ratings can be processed. The Writer class is used for synthesizing random data for creating simulations. Most importantly, the Hospital Advisor class sets up all of the other classes and would be the location of system outputs in simulations. The code can be found in the Appendix towards the end of the paper.

Patient Rating (PR) vs. Age

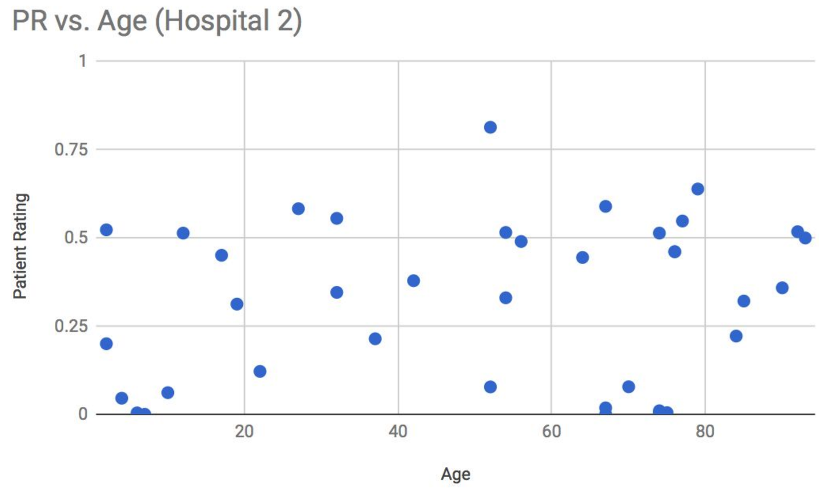
In the UCR formula, the relationship between PR and Age of patients for a certain hospital will give you the value ΔUCR_A. This value gives us an idea for which age groups get better treatment in each hospital. To find this value, every PR value for that hospital has to be averaged by age group. This gives us an average PR value for age group 0-9,10-19...90-99. Once these PR values are found, they are ranked from highest to lowest. The age group with the highest PR value is granted a ΔUCR_A value of 10, and the lowest is granted a value of 0. Now, the user specifies their age and the ΔUCR_A is whatever that age group's ΔUCR_A value is.

Here is an example for a 34 year old with two different examples:



Hospital 1	Avg PR	User-Compatibility Rating (ΔUCR_A)
Avg 0-9:	0.405	7
Avg 10-19:	0.276	5
Avg 20-29:	0.379	6
Avg 30-39:	0.175	4
Avg 40-49:	0.418	9
Avg 50-59:	0.412	8
Avg 60-69:	0.037	2
Avg 70-79:	0.041	3
Avg 80-89	0.443	10
Avg 90-99:	0.020	1

ΔUCR_A is 4 for the 34 year old's UCR for Hospital 1.



Hospital 2	Avg PR	ΔUCR_A
Avg 0-9:	0.155	1
Avg 10-19:	0.335	5
Avg 20-29:	0.352	6
Avg 30-39:	0.372	7
Avg 40-49:	0.379	8
Avg 50-59:	0.445	9
Avg 60-69:	0.263	2
Avg 70-79:	0.302	4
Avg 80-89	0.271	3
Avg 90-99:	0.459	10

ΔUCR_A is 7 for the 34 year old's UCR for Hospital 2.

Patient Rating (PR) vs. Gender

In the UCR formula, this relationship between Gender and PR gives us the value of CM. This CM value acts as a multiplier for the added delta values, and is used to give us an idea of which gender is treated better for each hospital. We made it a multiplier because it doesn't have the variety of score like our previous example. The best gender category only would grant you with two appropriate scores: 1 and 0. This is why, we use it as a multiplier. The way CM is calculated is the average PR is calculated for both genders at a hospital. Then, the user's gender is specified. If the user's gender is female, the CM = the average PR value for females in that hospital, and the same goes for males. Here is an example for a female.

Gender	Hospital 3 - CM	Hospital 4 - CM
Female	0.365	0.392
Male	0.363	0.364

The CM is going to be 0.365 for a female's UCR in Hospital 3.
The CM is going to be 0.392 for a female's UCR in Hospital 3.

Patient Rating (PR) vs. Disease Category

In the UCR formula, the relationship between the disease category and PR is ΔUCR_{DC}. This value is made to see which diseases are treated better for each hospital. The way that this is calculated is similar to ΔUCR_A. The PR for every disease category is averaged and given a ranking. When the user's specific disease category is specified, the ΔUCR_{DC} equals the ranking of that disease. Here is an example of someone looking to get treated for Cardiovascular Disease.

Hospital 5	PR	ΔUCR_{DC}
Cancer	0.313	2
CVD	0.415	7
Diabetes	0.320	3
Infectious Diseases	0.260	1
Mental Disorders	0.465	8
Neurological Disorders	0.336	4
Respiratory Diseases	0.387	6
Substance Abuse	0.381	5

ΔUCR_{DC} is 7 for the CVD patient's UCR for Hospital 5.

Hospital 6	PR	ΔUCR_{DC}
Cancer	0.326	2
CVD	0.374	3
Diabetes	0.612	8
Infectious Diseases	0.462	6

Mental Disorders	0.385	4
Neurological Disorders	0.404	5
Respiratory Diseases	0.193	1
Substance Abuse	0.558	7

ΔUCR_{DC} is 3 for the CVD patient's UCR for Hospital 6.

Doctor Experience (DE) vs. Disease Category

In the UCR formula, the relationship between disease category and doctor experience is ΔUCR_{DE}. This time the doctor datasets are used for each hospital to compare the disease category that the doctor deals with vs. their experience. This value is used to give us an idea of which diseases are treated with doctors of more experience for each hospital. ΔUCR_{DE} is calculated the same way as ΔUCR_{DC} with the ranking system. Here is an example of someone looking to be treated for Cardiovascular Disease.

Hospital 7	Years of Experience	ΔUCR_{DE}
Cancer	18.0	8
CVD	15.3	5
Diabetes	N/A (No doctors)	0
Infectious diseases	17.0	7
Mental Disorders	16.0	6
Neurological Disorders	16.0	6
Respiratory Diseases	N/A	0
Substance Abuse	N/A	0

ΔUCR_{DE} is 5 for the CVD patient's UCR for Hospital 7.

Hospital 8	Years of Experience	ΔUCR_{DE}
Cancer	17.5	8
CVD	8.3	1
Diabetes	12.5	3
Infectious diseases	15.0	7
Mental Disorders	13.2	4
Neurological Disorders	14.2	6
Respiratory Diseases	9.0	2

Substance Abuse	13.8	5

ΔUCR_{DE} is 1 for the CVD patient's UCR for Hospital 8.

Strengths and Weaknesses

Part 1: Strengths

1. Viewing and Sorting of Data

With the coding and our approach we have used, our model has encoded within each patient's score multiple variables. By doing so, we have created an easy way to sort through and view data across different type of patients. Patients can easily select a hospital's patient score report for a specifics disease category, a specified disease, specific ages, and a certain gender or a combination of any of the mentioned variables in order to compare not only the overall hospital rank but also individual patient scores in each sub-factor. This concept led to the UCR (User Compatibility Rating) - the amount of factors we take into consideration allow for the viewing of datas in multiple perspectives. We manage to summarize a large amount of factors into various ratings that give a general idea for comparing among hospitals.

2. Versatility and Flexibility Of Data

Our model is designed in a manner that we can add even more factors to the patient data set. For example, we can add another factor such as the budget and income of the patient and compare this data point with the average prices of treatment in hospitals. The more factors we taken into consideration in our model, the more accurate our hospital ratings become.

3. Consideration of Hospital Error

We make sure that any errors made by the doctor or the hospital is accounted for with the Error (E) constant. Rarely are hospital treatments and doctors, however experienced he or she may be, are prone to mistakes. By accounting for these errors, we again improve the accuracy of our ratings.

Part 2: Weaknesses

1. Dependence on Risk Factor

The Risk factor is the ratio between death's and incidence in a given year is used to estimate the severity of a disease. This risk factor is being used with the assumption that it accurately estimates the severity of a disease. This is not a precise form of measuring severity but it does give a rough idea.

2. The Distribution of Weight Across Factors

How we weigh duration of stay compared to risk factor is done in a manner that may not accurately reflect the desires of patients. While we believe both factors are desired for patients, due to the subjective nature of desires, we were unsure how to weigh the factors. We tried various functions to compare how duration of stay affected the patient rating compared to how risk factor would with different diseases. By doing this we eventually decided upon functions, but there was no precise mathematical relationship we could find to compare the weight besides what we deemed appropriate.

3. Use of Inconsistent Data In Model

Because much of our model was collected from numerous sources, this error could affect the accuracy of our model. However, we assume that hospitals will provide us with more precise data if such programs were created for real use.

Conclusion and Improvements

Our goal was to create a system that gives a user various hospital rankings so that the user may identify and choose the best place of treatment for the individual. To accomplish this, we created multiple rating systems (PR, HR, UCR). These rating systems took into consideration various factors such as DOS, DP, capacity and others. Although subjective to each person, some factors are prioritized and weighted heavily, while others are prioritized less and weighted less heavily in our model.

The primary strength of our solution is that because we treat each patient as a “data set” with various factors, we create personalized hospital search results for each individual patient. However, given more time and a wider and more reliable access to the data, we can extend our model to include even more factors to improve accuracy of our rating systems.

User Memo

Dear User,

This memo aims to clarify how our rating system works in order to result in various rankings that determines the quality of hospitals. This memo also aims to clarify the meaning of different values that defines the patient rating, hospital rating, user compatibility rating.

This hospital ranking system allows you to identify the overall best hospitals. It also allows you to identify which hospital best suits you as an individual patient. First, we have collected the most updated data that is provided from legitimate hospitals. This data allows us to create numerous models. The first model with which we evaluate the quality of the hospital is the Patient Rating. This rating takes into consideration the following: the average duration of stay of patients, the risk factor of a disease and possible errors made by the doctor or hospital. We assume that lower average duration of stay is preferable to long duration of treatments because increased treatment price. The risk factor is a measure of how severe or serious a disease is. For example, a common cold will have a significantly lower risk factor than stomach cancer because people are more likely to survive a cold than cancer. These factors are run through our program to give you a Patient Rating value. This patient rating becomes useful when comparing different aspects of a hospital (explained later in this memo).

Secondly, after entering your personal information such as age, gender, and disease (if your disease is known), the overall hospital could also be found. Factors such as Doctor to patient, hospital capacity, doctor experience, patient reviews of hospitals and the patient rating (explained above) are taken into consideration to find a overall hospital rating. Note that this is only the general hospital rating, so although a certain hospitals may be the highest ranked, it may not be the best suited place of treatment for you.

Because one general hospital rating is not sufficient to determine your ideal place of treatment, our program allows you to compare multiple aspects of multiple hospitals through the aforementioned patient rating. For example, you can compare and see in which hospital has the highest (best) patient ratings in which age group. You will be able to compare patient ratings to your gender, disease category in the same manner. Such comparisons will give you the User Compatibility Rating, a measure of how much a hospital and its records fit you personally. Just like the patient rating and hospital rating, a higher user compatibility rating indicates a more ideal location of treatment - this means that we a certain hospital might suit you very well but suit others very poorly.

All in all, our program gives you three primary values: the patient rating, hospital rating and user compatibility rating. Our program allows you to personalize your hospital search results so that you can find the hospital that suits you and your needs.

Appendix (Code + Simulated Data)

<u>HospitalAdvisor Class</u>

```
ArrayList<Hospital> hospitalList = new ArrayList<Hospital>();
void setup(){

 int hospitals = 2;
 int patientsL = 10;
 int patientsH = 200;

 String[] db = loadStrings("Database.txt");
 Writer writer = new Writer(hospitals,patientsL,patientsH,db);
 writer.write();
 writer.writeD();

 for(int i = 0; i < hospitals; i++){
  String[] hospital1 = loadStrings("hospitals/Hospital"+i+".txt");
  String[] doctors = loadStrings("doctors/Hospital"+i+".txt");

  int cap = round(random(200,400));

  hospitalList.add(new Hospital(hospital1,cap,db,doctors));
 }

 int maxLikes = 0;
 int minLikes = 250000;
 for(int i = 0; i < hospitalList.size(); i++){
  if(hospitalList.get(i).getLikes() > maxLikes) maxLikes = hospitalList.get(i).getLikes();
  if(hospitalList.get(i).getLikes() < minLikes) minLikes = hospitalList.get(i).getLikes();
 }

 for(int i = 0; i < hospitalList.size(); i++){
  float dp = hospitalList.get(i).getDP();
  float c = hospitalList.get(i).getC();
  float ade = hospitalList.get(i).getAvgExp();
  float de = (1.0/float(writer.getMax() - writer.getMin()))*(ade - writer.getMin());
  float pr = hospitalList.get(i).getAveragePR();
  float ur = (1.0/(float(maxLikes) - float(minLikes)))*(float(hospitalList.get(i).getLikes())-minLikes);
  float hr = (0.1*c) + (0.2*dp) + (0.2*de) + (0.3*pr) + (0.2*ur);
  hospitalList.get(i).setHR(hr);
  hospitalList.get(i).setDE(de);
```

```
  }

  //This part of the code depends on the data you are trying to sort
  PrintWriter output = createWriter("data.txt");
  for(int i = 0; i < hospitalList.size(); i++){
    output.println("Hospital"+(i+1));
    output.println("Hospital Rating,"+hospitalList.get(i).getHR());
    output.println("Category,Patient Rating");
    for(int j = 0; j < hospitalList.get(i).getPatientCount(); j++){

output.println(hospitalList.get(i).getPatient(j).getCategory()+","+hospitalList.get(i).getPatient(j).getRatin
g());
    }
  }
  output.flush();

}
```

Hospital Class

```
class Hospital{

  ArrayList<Patient> patientList = new ArrayList<Patient>(); //list of patients
  ArrayList<Doctor> doctorList = new ArrayList<Doctor>();

  int capacity;
  String[] db;

  float dp;
  float c;
  float de;
  float hr;

  int ur;

  Hospital(String[] data, int capacity,String[] db,String[] doctors){
    for(int i = 0; i < data.length; i++){
      String[] line;
      line = split(data[i],",");
      String name = line[0];
      String age = line[1];
      boolean gender;
      if(line[2].equals("0")){
        gender = false;
      }else{
        gender = true;
```

```
    }
    String riskFactor = line[3];
    String days = line[4];
    boolean discharge;
    if(line[5].equals("0")){
      discharge = false;
    }else{
      discharge = true;
    }

    String category = line[6];
    String disease = line[7];

    boolean error = false;
    if(line[8].equals("1")){
      error = true;
    }

    boolean read = false;
    if(line[9].equals("1")){
      read = true;
    }

    patientList.add(new
Patient(name,int(age),gender,float(riskFactor),int(days),discharge,category,disease,db,error,read));
  }

  this.capacity = capacity;
  this.db = db;

  for(int i = 0; i < doctors.length; i++){
    String line[] = split(doctors[i],",");
    String name = line[0];
    String cat = line[1];
    int years = int(line[2]);
    doctorList.add(new Doctor(name,cat,years));
  }

  dp = float(doctorList.size()) / float(patientList.size());
  c = 1 - (float(patientList.size()) / float(capacity));
  ur = floor(random(0,250000));

 }

 Patient getPatient(int index){
  return patientList.get(index);
 }
```

```
Doctor getDoctor(int index){
  return doctorList.get(index);
}

float getAveragePR(){
  float sum = 0;

  for(int i = 0; i < patientList.size(); i++){
    sum += patientList.get(i).getRating();
  }

  float average = sum / patientList.size();

  return average;
}

float getMedianPR(){
  int l = patientList.size();
  float[] pl = new float[l];
  for(int i = 0; i < l; i++){
    pl[i] = patientList.get(i).getRating();
  }
  pl = sort(pl);

  float median = 0;
  if(l % 2 == 0){
    median = (pl[(l/2)] + pl[(l/2)-1])/2;
  }else{
    median = pl[(l/2)+1];
  }
  return median;
}

float getSDPR(){
  int l = patientList.size();
  float sum = 0;
  for(int i = 0; i < l; i++){
    sum += (patientList.get(i).getRating() - this.getAveragePR());
  }
  float sd = sqrt(sum / (l-1));
  return sd;
}

int getPatientCount(){
  return patientList.size();
}

int getDoctorCount(){
```

```
  return doctorList.size();
 }


 float getDP(){
  return dp;
 }

 float getC(){
  return c;
 }

 float getAvgExp(){
  float sum = 0;
  for(int i = 0; i < doctorList.size(); i++){
   sum += doctorList.get(i).getYears();
  }
  return sum / doctorList.size();
 }

 int getLikes(){
  return ur;
 }

 void setHR(float hr){
  this.hr = hr;
 }

 float getHR(){
  return hr;
 }

 void setDE(float de){
  this.de = de;
 }

}
```

<u>Patient Class</u>

```
class Patient {

 String name; // name of patient : Patient#1, Patient#1
 int age; // age of patient
 boolean gender; // gender of patient: 0-Male 1-Female
 float riskFactor; // risk factor of patient
```

```
 int days; // duration of stay of patient
 boolean discharge; // dead or discharge 0-Dead 1-Discharged
 String category;
 String disease;
 String[] db; // database of sicknesses, risk factors and d.o.s
 boolean error; // if there is error 0-No 1-Yes
 boolean read; //if readmitted due to error

 Patient(String name, int age, boolean gender, float riskFactor, int days, boolean discharge, String
category, String disease,String[] db,boolean error,boolean read){
   this.name = name;
   this.age = age;
   this.gender = gender;
   this.riskFactor = riskFactor;
   this.days = days;
   this.discharge = discharge;
   this.category = category;
   this.disease = disease;
   this.db = db;
   this.error = error;
   this.read = read;
 }


 float getRating(){
   if(error && discharge) return .25;
   if(error && !discharge) return .01;

   float rating = .5;
   if(read) rating = .25;

   float dRating;
   if(discharge){
     dRating = pow(1.5,riskFactor/100.0) - 1;
     //System.out.println(dRating+"DIS");
   }else{
     dRating = -.5/pow(10,riskFactor/100);
     //System.out.println(dRating+"DEATH");
   }

   rating += dRating;

   float avgDos = 0.0;
   for(int i = 0; i < db.length; i++){
     String[] line = split(db[i],",");
     if(line[1].equals(disease)){
       avgDos = float(line[3]);
     }
```

```
    }

    //System.out.println(rating);

    dRating = 0;
    if(discharge){
      dRating = -(1/(4*pow(avgDos,3)))*pow((days-avgDos),3);
    }else{
      dRating = (1/(4*avgDos*avgDos*avgDos))*pow((days-avgDos),3);
    }

    rating += dRating;

    if(rating < 0) return 0;

    return rating;
  }

  String toString(){
    return name + "," + age + "," + gender + "," + riskFactor + "," + days + "," + discharge +
","+category+","+disease+","+this.getRating();
  }

  int getAge(){
    return age;
  }

  String getGender(){
    if(gender){
      return "Female";
    }else{
      return "Male";
    }
  }

  String getCategory(){
    return category;
  }

}
```

Doctor Class

```
class Doctor{

  String name;
```

```
  String category;
  int years;

  Doctor(String name, String category, int years){
    this.name = name;
    this.category = category;
    this.years = years;
  }

  int getYears(){
    return years;
  }

  String getCategory(){
    return category;
  }

}
```

Writer Class

```
class Writer {

  int hospitals; // of hospitals files
  int patientsL; // of patients in each hospital
  int patientsH; // of patients in each hospital
  String[] db; // database of all risk factors (Category,Disease,RF)

  int min = 0;
  int max = 30;

  //constructor method
  Writer(int hospitals, int patientsL, int patientsH ,String[] db){
    this.hospitals = hospitals;
    this.patientsL = patientsL;
    this.patientsH = patientsH;
    this.db = db;
  }

  void write(){
    for(int i = 0; i < hospitals; i++){
      PrintWriter output = createWriter("hospitals/Hospital"+i+".txt");
      float patients = random(patientsL,patientsH);
      for(int j = 0; j < patients; j++){
        String[] randLine;
        int randIndex = floor(random(0,db.length));
```

```
      String line = db[randIndex];
      randLine = split(line,",");
      String name = "Patient"+j;
      int age = floor(random(1,95));
      int gender = round(random(0,1));
      float riskFactor = float(randLine[2]);
      float dos = float(randLine[3]);
      int days = ceil(random(0.3,2)*dos);
      int discharge = round(random(0,1));
      String category = randLine[0];
      String disease = randLine[1];

      float chanceError = random(0,1);
      float readError = random(0,1);
      int error = 0;
      int relapse = 0;
      if(chanceError > .98){
        error = 1;
      }
      if(readError > .99){
        relapse = 1;
      }

output.println(name+","+age+","+gender+","+riskFactor+","+days+","+discharge+","+category+","+dise
ase+","+error+","+relapse);
    }
    output.flush();
  }
 }

 void writeD(){
  for(int i = 0; i < hospitals; i++){
    PrintWriter output = createWriter("doctors/Hospital"+i+".txt");

    float patients = random(patientsL,patientsH);
    float doctors = ceil(patients * random(0.2,1.1));
    for(int j = 0; j < floor(doctors) ;j++){
      String name = "Doctor"+j;
      String[] line = split(db[floor(random(0,db.length))],",");
      String category = line[0];
      int years = floor(random(1,30));
      if(years > max) max = years;
      if(years < min) max = years;
      output.println(name+","+category+","+years+","+max+","+min);
    }
    output.flush();
  }
```

```
 }

 int getMin(){
   return min;
 }
 int getMax(){
   return max;
 }



}
```

Data

The data our team simulated is quite large to put on this document. Below is a shortened link where the data can be viewed and downloaded if needed.

https://bit.ly/2JJZ8vC

Sources

List of Disease Groups: KPHO
Available Disease Categories

http://www.who.int/cardiovascular_diseases/en/cvd_atlas_01_types.pdf
CVD List

http://www.who.int/healthinfo/global_burden_disease/estimates/en/index1.html
Global Numbers of Death from Diseases - 2015 - Data Download Link

http://www.who.int/healthinfo/global_burden_disease/GBD_report_2004update_part3.pdf
PDF - Incidence of Certain Diseases - 2004