Elephantiasis Analysis – Extended Report

Technical Report · December 2017

CITATIONS	READS
0	37

2 authors, including:

Akash Nagaraj
PES Institute of Technology
8 PUBLICATIONS **1** CITATION

Elephantiasis Analysis

UE15CS314 : Big Data | 5th December, 2017

SNo	Name
1	Mukund Sood
2	Akash Nagaraj
3	Bishesh Sinha

Introduction

Lymphatic filariasis, also known as elephantiasis, is a human disease caused by parasitic worms known as filarial worms. Most cases of the disease have no symptoms. Some people, however, develop a syndrome called elephantiasis, which is marked by severe swelling in the arms, legs, breasts, or genitals. The skin may become thicker as well, and the condition may become painful. The changes to the body may harm the affected person's social and economic situation.

The worms are spread by the bites of infected mosquitoes. Three types of worms are known to cause the disease: Wuchereria bancrofti, Brugia malayi, and Brugia timori, with Wuchereria bancrofti being the most common. These worms damage the lymphatic system. The disease is diagnosed by microscopic examination of blood collected during the night. The blood is typically examined as a smear after being stained with Giemsa stain. Testing the blood for antibodies against the disease may also permit diagnosis. Other roundworms from the same family are responsible for river blindness.

Prevention can be achieved by treating entire groups in which the disease exists, known as mass deworming. This is done every year for about six years, in an effort to rid a population of the disease entirely. Medications used include antiparasitics such as albendazole with ivermectin, or albendazole with diethylcarbamazine. The medications do not kill the adult worms but prevent further spread of the disease until the worms die on their own. Efforts to prevent mosquito bites are also recommended, including reducing the number of mosquitoes and promoting the use of bed nets.

Our main Objective was analyzing Integration of Modern Dermatology with Ayurveda in the treatment of Lymphatic Filariasis.

Provided with data from the Institute of Applied Dermatology regarding over 2400 patients, our main Objective was to analyse this data to find trends and patterns, to aid doctors in the treatments of this disease. The data provided us with information regarding the effectiveness of a wide range of medicines including Ayurvedic and Allopathic medicines. We were also provided with data regarding the amount of time a patient spent doing Yoga, as recommended by the treating doctor.

Using this data we applied a wide range of algorithms to find trends in the data, to get a better understanding into better and more streamlined treatments tracks based on the patient. We were able to do this by performing various analyses, to find which medicine works best, or how patients from different subgroups of a population react to particular medicines.

Related work

1. *Narahari, Saravu R., Terence J. Ryan, Kuthaje S. Bose, Kodimoolon each therapeutic system and a working knowledge of approach to their clinical diagnosis.e S. Prasanna, and Guruprasad M. Aggithaya. "Integrating modern dermatology and Ayurveda in the treatment of vitiligo and lymphedema in India."*

A team of doctors from modern dermatology, Ayurveda, yoga therapy, and homeopathy studied recruited patients to develop mutual orientation

This paper talks about the various techniques as well as the combinations of techniques used by modern medicine to treat difficult cases of diseases like Filarial Lymphodema.

2. *Narahari, Saravu R. "Collaboration culture in medicine."*

Integrative Medicine selects best practice of public health and biomedicine. IM protocols in India. The major knowledge base of allopathy and Complementary and alternative systems of medicine is derived from the same human observation of natural events.

This paper talks about the different methods of biomedicine as well as Complementary and alternative systems of medicine used in India, as well as the IM protocols. The author is affiliated to IAD, and the knowledge-base is derived from natural events and real-world observations.

3. *Narahari, Saravu R., Kodimoole S. Prasanna, and Kandathu V. Sushma. "Evidence-based integrative dermatology."*

Studies conducted at the IAD have created a framework for evidence-based and integrative dermatology (ID). This paper gives an overview of the advances in ID with an example of *Lichen planus*.

This paper provides a clear understanding of the kind of analysis to be performed by us, to achieve evidence-based results, as it works around an example.

● Algorithm/Design

Over the course of this project, we used multiple algorithms, to analyze the data and come up with evidence-based results. A few of the algorithms we have used:

- ● ANOVA - Analysis of Variance

  Analysis of variance (ANOVA) is a collection of statistical models and their associated procedures used to analyze the differences among group means.

- ● Student's t- tests

  The t-test is any statistical hypothesis test in which the test statistic follows a Student's t-distribution under the null hypothesis.

  A t-test is most commonly applied when the test statistic would follow a normal distribution if the value of a scaling term in the test statistic were known. When the scaling term is unknown and is replaced by an estimate based on the data, the test statistics follow a Student's t distribution. The t-test was used, to determine if two sets of data are significantly different from each other.

- ● Chi-Square tests

  A chi-squared test, also written as test, is any statistical hypothesis test wherein the sampling distribution of the test statistic is a chi-squared distribution when the null hypothesis is true.

  In our case, the test was applied when you have two categorical variables from a single population. It is used to determine whether there is a significant association between the two variables.

- ● RMANOVA - Repeated Measures Analysis of Variance

  Repeated measures design uses the same subjects with every branch of research, including the control. For instance, repeated measurements are collected in a longitudinal study in which change over time is assessed.

- ● Market Basket Analysis

  Although this may seem a little strange, we thought that the best algorithm to analyze combinations of the treatments prescribed would be to use Market Basket analysis, which measures the support and confidence of individual rules that are automatically generated by the algorithm. This algorithm, commonly used in the e-commerce ecosystem proved to be very successful and returned great results.

Over the course of this project, we followed the tasks set by a senior statistician and executed as many as we could with the data we had. A few of the ideas did not have enough data, or in the case of a few others, the data had far too many discrepancies to work with and attain useful insights.

The general procedure we followed for most of the tasks is as follows:
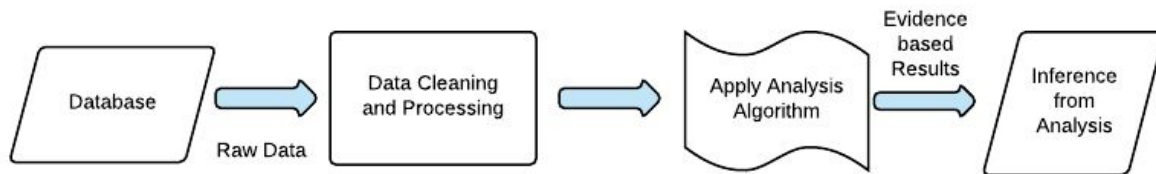
1. Extract the data in consideration
2. Clean the data, and do the required preprocessing (converting textual data into categorical data we can work with, etc.)
3. Analyze the data atomically and pick the best algorithm to perform a particular analysis on it.
4. Inferences from the result of the algorithms
5. Report the result and draw conclusions

Experimental Results:

Over the course of this project, we followed a basic guideline set by a senior statistician. Although some of the analyses in this guideline was not feasible, we attempted to work around it wherever possible, but, we could not do much in cases with a lack of data, or unreadable data.

When we began, the data seemed to be full of discrepancies and had no formal documentation to understand the meaning of each table, columns of tables, and even the individual values themselves were hard to understand. This inspired us to document the tables, and individual columns to aid teams that would work on the same data.

Overall, we worked on all of the analyses suggested, and reported the ones that could not be done, along with the reason.

Analysis 1 - Volume Decrease vs Follow Up Number: strong correlation between follow up number and the volume decrease

Analysis 2 - Impact of C-Bandage Stocking: the average change at follow up 1 of patients who used C-Bandage Stocking is 2.179 units whereas those who did not use it was 1.472 units, thereby showing that it was an effective treatment.


Change in Volume vs Followups for Patient
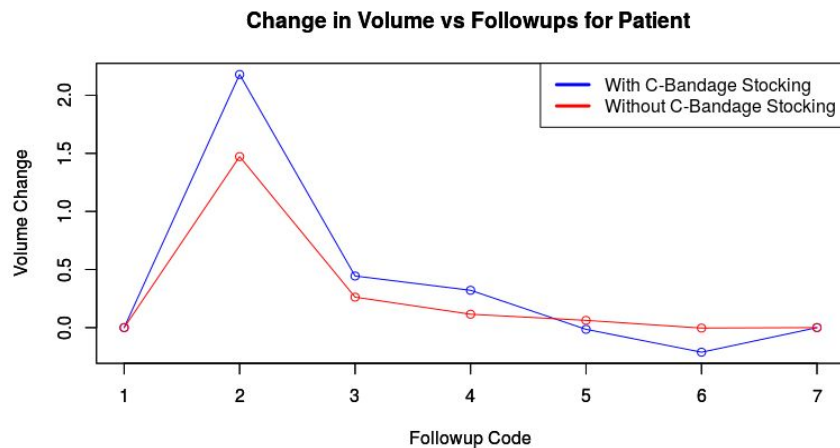
Analysis 3 - Impact of Compression: patients who did not use compression showed a better change over those who used compression. However, upon closer inspection, the number of patients who used compression were 4.5 times the number of patients who did not use compression, so we could attribute this to a lack of data.

Analysis 4 - Impact of Wrapping Cloth: The results show that those who used wrapping cloths faired marginally better than those who did not.

Change in Volume vs Followups for Patient (Normalized)
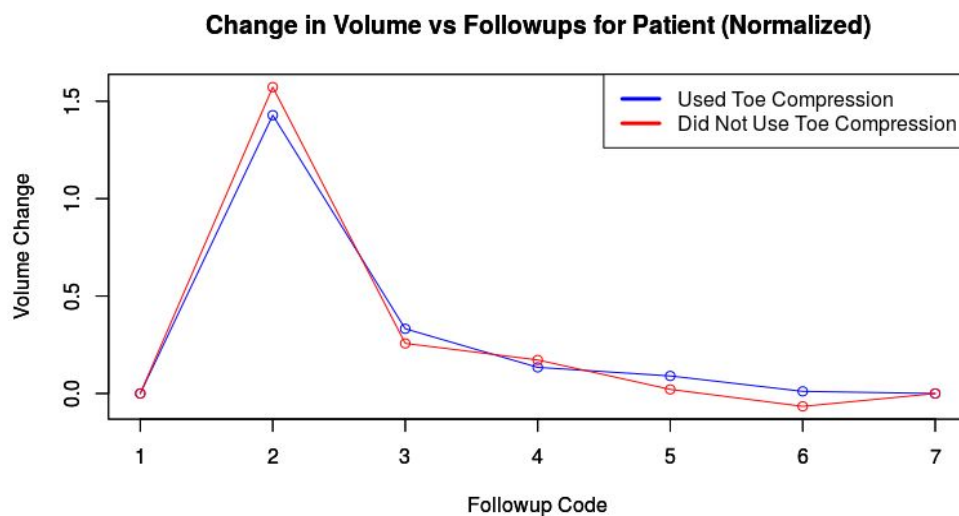
Analysis 5 - Impact of Toe Compression: patients who used Toe Compression faired marginally better than those who did not.

Analysis 6 - Impact of Sterilization: results show a rather counterintuitive relation, those who did not sterilize their wound faired better than those who did, but again, we see that the number of patients that did not sterilise were 1/10th of the patients that did, so we can attribute this to a lack of data.

Change in Volume vs Followups for Patient (Normalized)
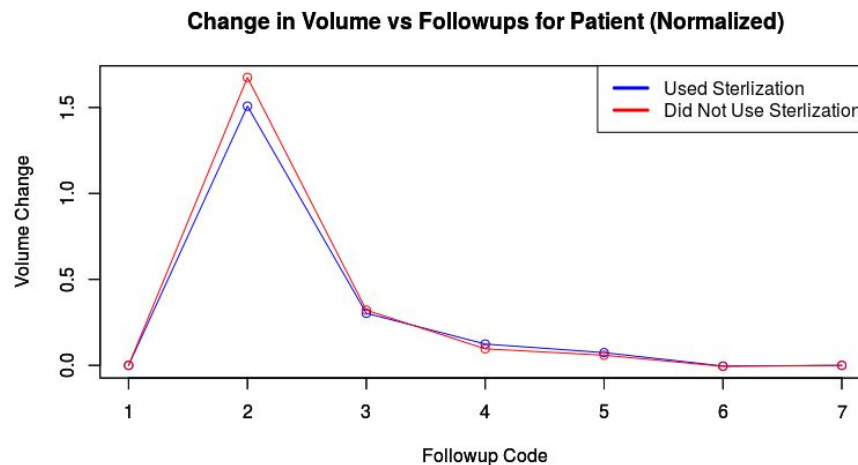
Analysis 7 - Impact of Sponge Mould Folds: The results show that those who used Sponge Mould Folds faired better than those who did not and this would be an advised course of action.

```
> inspect(rules5)
     lhs                               rhs                                                  support   confidence lift       count
[1] {vol_change=Excellent_Change} => {using__mould=Did_Not_Use_Mould}                       0.1151392 0.5000000  0.9873998  397
[2] {vol_change=Excellent_Change} => {toe_compression=Used_Toe_Compression}                 0.1424014 0.6183879  1.0834357  491
[3] {vol_change=Excellent_Change} => {spong_mould_folds=Used_Sponge_Mould_Folds}            0.1548724 0.6725441  1.1400846  534
[4] {vol_change=Excellent_Change} => {changing_cleaning=Not_Available}                       0.2033063 0.8828715  1.0595688  701
[5] {vol_change=Excellent_Change} => {compression_practice=Used_Compression}                0.2033063 0.8828715  1.0580956  701
[6] {vol_change=Excellent_Change} => {sterilization=Used_Sterilization}                     0.1980858 0.8602015  1.0269996  683
[7] {vol_change=Excellent_Change} => {wrapping_cloth=Used_Wrapping_Cloth}                   0.2073666 0.9005038  1.0436763  715
[8] {vol_change=Excellent_Change} => {c_bandage_stocking=Not_Available}                     0.2064965 0.8967254  0.9656182  712
```

Analysis 8 - Analysis of Combinations of Treatments: for this, we made an attempt at Market Basket Analysis. We saw that in all cases, the best result is obtained by *Using a Wrapping Cloth*, *Sterilizing the Wound* and *Using Plain Compression*, and we can say this with a Confidence of over 0.80 in all cases.

Analysis 9 - Impact of Yoga before using Medicine:
<u>Null Hypothesis -</u> Change in volume is independent of doing Yoga before the use of any medicine
<u>Alternate Hypothesis -</u> The two variables are not independent.
The P-value obtained is a very low 0.05361, and as the threshold we have taken is 0.05, we cannot reject the Alternative Hypothesis, and conclude that there is an effect of doing yoga and a change in volume.
Based on the results we can see a strong relation between change in volume and followups, but a very weak relation between doing Yoga and a change in volume.

Analysis 10 - Impact of Yoga after using Medicine:

Null Hypothesis - Change in volume is independent of doing Yoga after the use of any medicine
Alternate Hypothesis - The two variables are not independent.

```
> chisq.test(post_yoga)

        Pearson's Chi-squared test

data:  post_yoga
X-squared = 20.969, df = 8, p-value = 0.007229
```

```
> anova_post
Analysis of Deviance Table (Type II tests)

Response: vol_change
                        Df    Chisq Pr(>Chisq)
time_post_yoga           2   2.1724     0.3375
followup_no              1 619.5296     <2e-16 ***
time_post_yoga:followup_no 2   0.0172     0.9915
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The P-value is a very low 0.007229, and as the threshold we have taken is 0.05, we can say the doing Yoga after taking the medicine does not make a difference.
Based on the results below we can see a strong relation between change in volume and followups, but a very weak relation between doing Yoga and a change in volume.

Analysis 11 - Impact of Previous Ailments: We ran an analysis on how people affected by diseases (Diabetes, Hypertension, Cardiac Disease, Liver Disease, Renal Disease, Orthopaedic Problem) in the past react to being treated for filariasis, and we see that patients who suffered from Liver Disease show a significant uptake as time goes by, while patients with Orthopaedic problems show the least positive changes, so it could be said that additional care must be given to them.
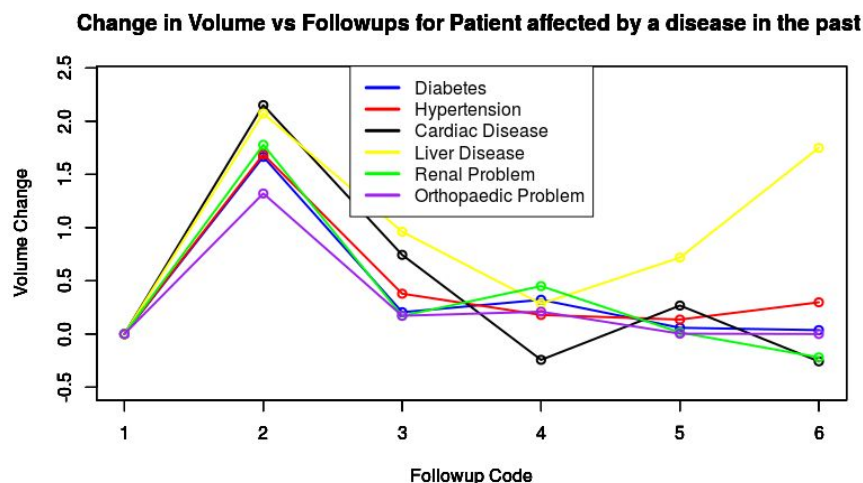

Analysis 12 - Effect of Varicose Veins: Varicose Veins do not appear to a very strong factor in determining which patients catch the disease, and it could possibly be pointing towards it being a different factor being the main cause in Elephantiasis

Analysis 13 - District wise Distribution of Patients: From the analysis of change in volume per follow up against the district for the top 5 districts, we can see that Bangalore clearly outperforms the other districts, and is much better than Gulbarga, which is also an inland district. Although, initially, there is a huge difference between Gulbarga, and the other coastal districts. But, again, we can't say much from this, owing to the fact that Bangalore starts off worse than a few coastal districts, but ends up with the best result. This could be attributed to a hidden factor.

District Wise Change in Volume vs Followups

Distribution of Patients
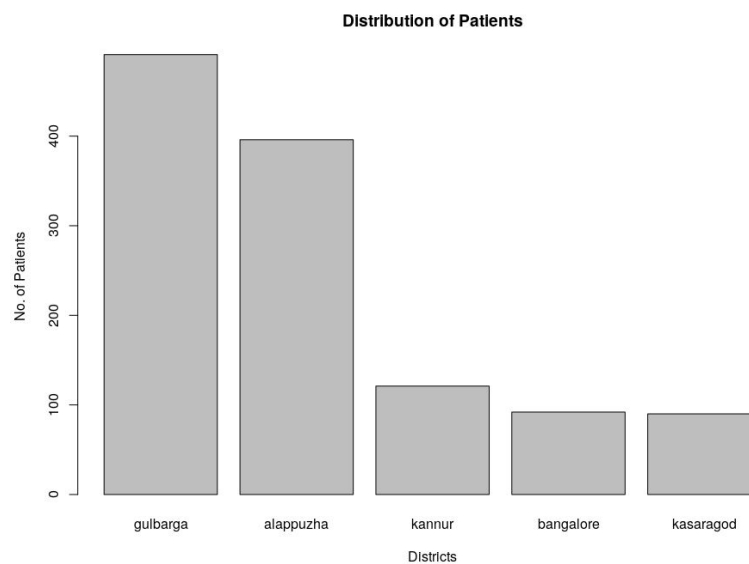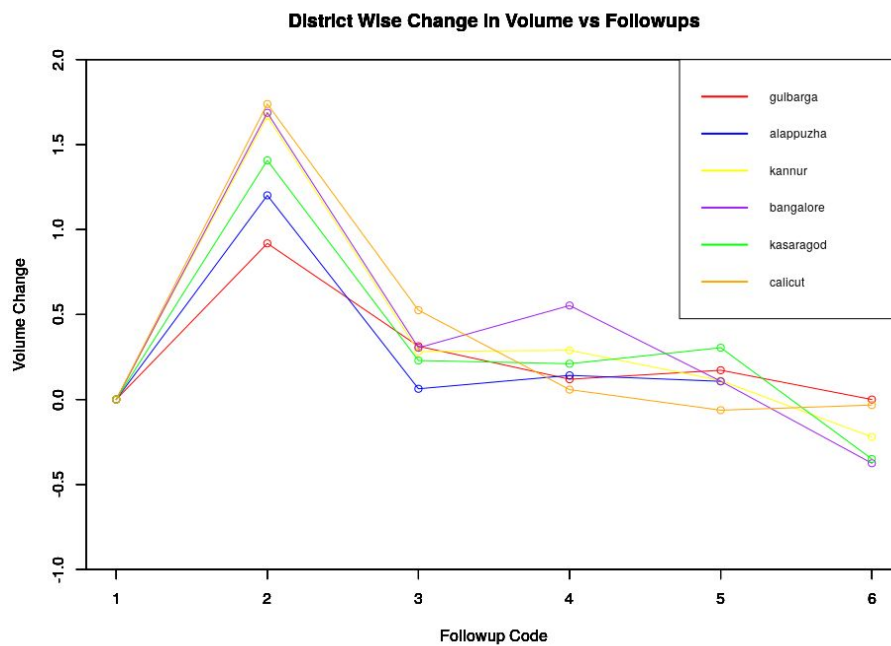
Analysis 14 - Region-wise Distribution of Patients: Patients from India, make up about 98.7% of the all the patients IAD treats, while there are merely 1 or 2 patients from the other countries.

Analysis 15 - Impact of Age: Surprisingly, it was found that the age does not appear to a very strong factor in catching the disease, or even recovery.
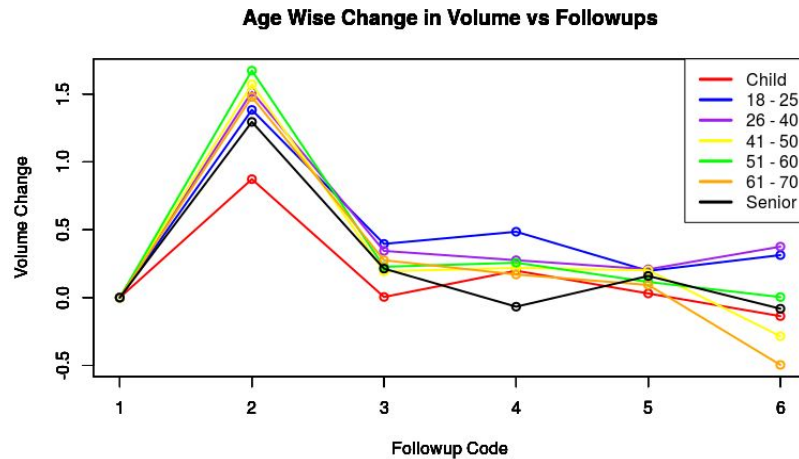
Analysis 16 - State wise Distribution of People: It was found that there is a high correlation between the state a person belongs to and the limb volume. This could be attributed again, to a hidden factor.
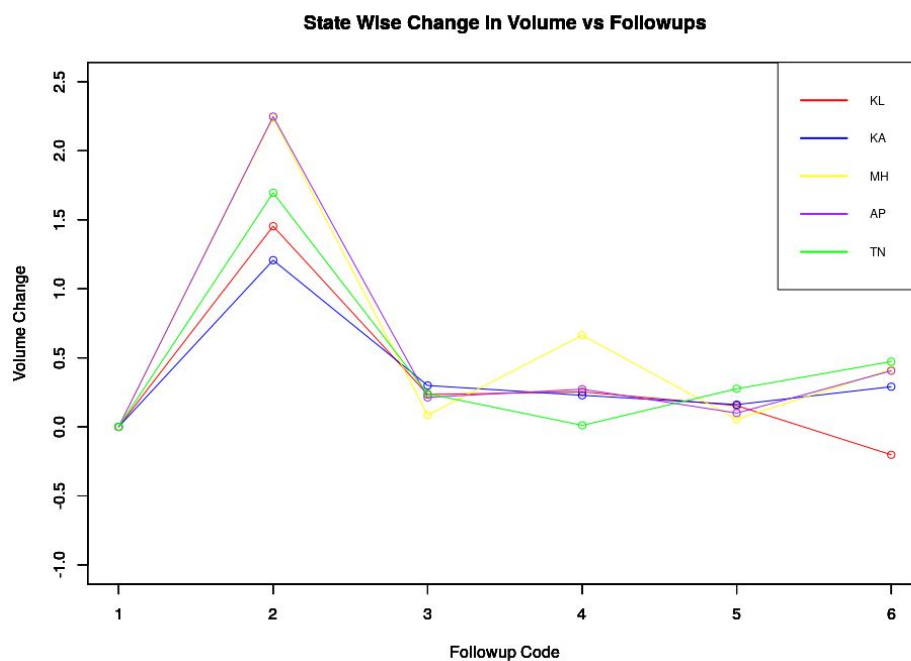
Analysis 17 - Impact of Gender: Upon running ANOVA, we see that the gender does strongly impact the patient's limb volume through followups.

Analysis 18 - Impact of Socioeconomic Status: Again, we see that there is a strong relation between the socioeconomic status of a patient and his/her recovery from the disease.

Socio-Economic Status Wise Change in Volume vs Followups
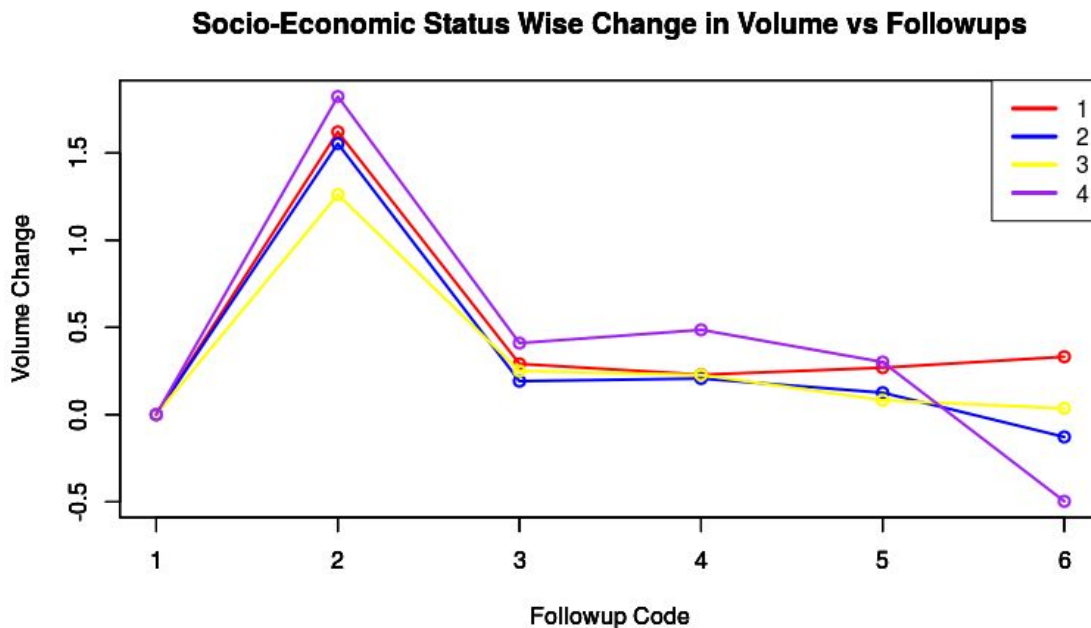
Analysis 19 - Distribution of Ration Card Type: Although there was no real data about how Ration Card Type 1 differs from Ration Card Type 2, we found that patients with Ration Card Type 1 recovery faster that patients with Ration Card Type 2.

Analysis 20 - Impact of Family Income: We see that patients earning an income of Level 4 show the best response, and that recovery is highly dependent on family income.

Analysis 21 - Impact of Employment Status: Patients who are part of Employment Status 1, 2 and 3 are worst affected by Elephantiasis, there is also a strong relation between the employment status and the limb volume

Analysis 22 - Impact of Ulcer Growth: From the analysis, we see that patients who suffered with Ulcers were slightly worse off than those that did not.

	Ointment.Type	Times.Used	Times.Effective	Effectiveness
1	Triben	1779	1225	68.85891
2	Rasnadi	677	479	70.75332
3	Loceryl	53	44	83.01887
4	Candid	72	59	81.94444
5	Tricawin	171	128	74.85380
6	Keto Based	111	77	69.36937
7	Others	537	381	70.94972

Analysis 23 - Effectiveness of Intertrigo Ointments: We see that Loceryl had the highest effectiveness and had the least usage, whereas, Triben was used the most but consequently had the lowest Effectiveness, however this cannot be taken as a sound representation because of the disproportionate use of all the Ointments, more data, and an equitable use of all ointments would give us a much better chance at analysis, hence we would still continue to prescribe Triben or Rasnadi.

Swelling Duration wise Change in Volume vs Followups
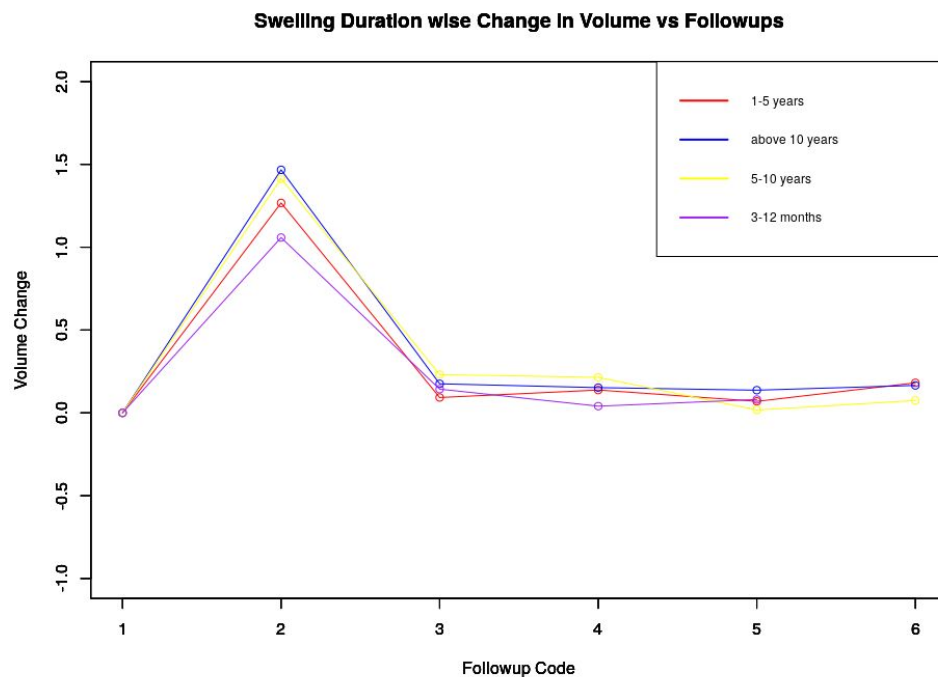
Analysis 24 - Distribution of Swelling Duration: We can see that the swelling is typically initially severe in patients who have the swelling for longer periods of time, and that there is a direct correlation between the average swelling at the first follow up and the overall duration of the swelling.

Analysis 25 - BEP vs Combination of Yoga and IMLD: Not possible due to lack of data.

Analysis 26 - IE Frequency vs Yoga and IMLD: The results show that those who used Sponge Mould Folds faired better than those who did not and this would be an advised course of action.

Analysis 27 - IE Frequency and only IMLD: The results show that those who used Sponge Mould Folds faired better than those who did not and this would be an advised course of action.

Analysis 28 - IE Frequency vs Neither Yoga nor IMLD: The results show that those who used Sponge Mould Folds faired better than those who did not and this would be an advised course of action.

set1	0 obs. of 31 variables	
set2	0 obs. of 97 variables	
set3	0 obs. of 97 variables	
set4	0 obs. of 97 variables	

set1	49 obs. of 31 variables	
set10	10 obs. of 67 variables	
set11	5 obs. of 67 variables	
set12	12 obs. of 67 variables	
set2	10 obs. of 67 variables	
set3	5 obs. of 67 variables	
set4	12 obs. of 67 variables	
set5	2955 obs. of 14 variables	
set6	10 obs. of 67 variables	
set7	5 obs. of 67 variables	
set8	12 obs. of 67 variables	
set9	774 obs. of 31 variables	

Future Enhancements:

Working with real-world data has proven to be quite a strenuous task, and to take this project further, a lot more data must be collected. When the data becomes too large to handle locally, we believe the best way to handle it would be to port the entire data to HIVE, and work on parallelizing work across the large data using Map-Reduce algorithms to attain analyses similar what we have produced using R.

We could also provide/build an interface to make data collection easier, and also provide real-time analyses and results. For example, if a patient were to come into a clinic seeking treatment, the doctor could enter data related to the patient into the system that we build and get a real time analysis of the patient's background, initial volume and a host of other parameters to offer the best course of treatment. This coupled with the domain knowledge that the doctor already possesses, could further aid in the cure of this disease. As the amount of data collected grows we can possible scale up and port the data to HIVE as stated above.

We also hope to publish our result and hope to influence IAD to amend their techniques to treat Elephantiasis more efficiently, with our evidence based results.

References

➔ Narahari, Saravu R., Terence J. Ryan, Kuthaje S. Bose, Kodimoole S. Prasanna, and Guruprasad M. Aggithaya. "Integrating modern dermatology and Ayurveda in the treatment of vitiligo and lymphedema in India." *International journal of dermatology* 50, no. 3 (2011): 310-334.
➔ Narahari, Saravu R. "Collaboration culture in medicine." *Indian journal of dermatology* 58, no. 2 (2013): 124.
➔ Narahari, Saravu R., Kodimoole S. Prasanna, and Kandathu V. Sushma. "Evidence-based integrative dermatology." Indian journal of dermatology 58, no. 2 (2013): 127.
➔ https://www.r-bloggers.com - For references to help with R