**Progress report – Jan 2019**

**Updated title:** Analysis of hospital based ayurvedic clinical practice to gain real world data knowledge

**Old Title:** Observational analysis of ayurvedic principles, ayurvedic hospital data, and patient outcomes

**Change of guide:**

1. Dr. Girish Tillu has taken up a full time position in Savitribai Phule Pune University. Hence he cannot continue being a guide.
2. Dr. Ashwini Godbole who has been on the DAC so far has agreed upon taking the role of a Primary Guide.
3. Dr. Girish Tillu will continue in the role of a “Co-guide” as well as will be a part of DAC going forward. Dr. Ashwini Mathur will continue in the role of a “Co-guide” and will be a part of DAC as well.

By Vinay Mahajan, Girish Tillu, Ashwini Mathur, Ashwini Godbole

Summary: The following progress has been made so far between July 2018 and Jan 2019

Presentation at 8th World Ayurveda conference, Ahmedabad Dec 2018: I was invited to present the real world data science view and patterns in data in transdisciplinary session for ayurveda. The presentation was based on the data analysis carried out and additional scientific material from Dr. Girish Tillu and Dr. Ashwini Mathur.

The chapter wise split of the thesis document will be as follows (subject to change in future based on discussions and ongoing research):

Chapter 1: Retrospective analysis of published literature, this chapter will provide empirical evidence about the status of Ayurvedic clinical research.

Chapter 2: Parallels between the ICH guidelines and Classical Samhitas to form the underlying fundamental basis of scientific principles. Perspectives on data science relevant to the research question will be demonstrated.

Chapter 3: Description about the hospital data, 5 W’s of data. Who, Why, What, Where, When and How.

Chapter 4: Description of software and analytical tools used, statistical analysis methods employed, data analysis and visualization method framework

Chapter 5: 2 to 3 specific disease areas to be used for detailed analysis.

The following section outlines the analysis carried out so far:

# Analysis carried out using Tableau, R, Java D3 and SPMF (Sequential Pattern Mining) Java library

This document covers the details of analysis carried out on the Hospital database.  
The Source data used is the Hospital database SQL programming and R programming is used to create analysis datasets. These analysis datasets are then used as inputs to Tableau for generating various interactive visual displays

## Displays created in Tableau [<https://public.tableau.com/profile/frlht#!/>]

**Based on data from SQL database till Oct 2017**

## 1. Viz name: 01SQL\_Dis\_Med\_Ser

| **Sheet name** | **Description** |
| --- | --- |
| RMSD\_Met\_patients | Frequency table by gender and high level disease classification, there are more number of RMSD patients compared to the Metabolic, Metabolic and RMSD patients. There are more number of female RMSD patients compared to males. There are similar number of males and females in Metabolic disease categories |
| Visit\_Duration | Total duration of hopsital visits is calculated as the maximum date of hopsital visit - minimum date of hospital visit + 1 in days for each patient. This duration is plotted by each disease category.  - The median duration for RMSD was more than that for other categories Median duration by disease category for females / males: Metabolic: 325.5 / 408 Metabolic and RMSD: 530.5 / 559.5 RMSD: 690 / 631 |
| Patient\_Visit\_View | This is a listing of individual patient by disease categories. The x-axis displays study day going from day 1 to last visit for each patient. Each bar represents a single study day. The IP visits are marked in Blue and OP visits are marked in Orange colour. For each visit, what kind disease type has been reported is displayed. The tooltip provides additional information related to Total duration of visits to hospital, description of disease, medicine name, first (minimum) day on which Metaolic disease was reported, first (minimum) day on which RMSD disease was reported. - Patients suffering from RMSD type of diseases have more frequent visits to the hospital than those suffering from Metabolic diseases. |
| 1stDay\_Met\_Disease | Summary statistics by gender for first (minimum) day on which Metaolic disease was reported. |
| DisType\_Diseases | Frequency table for individual diseases by disease category and gender Prameha, Madhumeha and Sthaulya are top 3 most frequently reported diseases. - Prameha and Madhumeha are reported more by males than females. - Vaatavyaadhi - Sandhigata Vaata, Vaatavydahi, Vaatavydahi - Gridhrasee, Sthanabhedena Shoola - Katee Shoola and Sthanabhedena Shoola - Katee Graha are top 5 reported diseases in RMSD. - Vaatavyaadhi - Sandhigata Vaata, Amavaata, Vaatavyaadhi - Asthigata Vaata, Vaatavyaadhi - Vaatakantaka are reported more by females compared to males. |
| MedType\_DisType | Frequency table for medicine categories by gender. Frequency counts for each prescribed medicine is reported. - The source variable reported in the database needs to be modified to seperate the medicine name, the quantiy and the manufactures name. - Current frequency counts show a 1 medicine prescibed to more than 3000 patients, 2 medicines to more than 2000 patients, 19 medicines to more than 1000 patients |
| Medicine\_DisCode | Frequency table for disease categories, Medicine by gender and individual disease. - as explained earlier, the medicine name variable needs to be modified to get an accurate picture of prescribed medicines |
| DiseaseByStudyDay | Frequency counts of patients reporting a disease on a particular study day under each disease category. - Most of the diseases have the highest frequency reported on day 1 and a steep drop in frequency counts is observed. |
| DiseaseByStudyDay - by season | Analysis similar to DiseaseByStudyDay, the frequency counts are reported by rutus |
| MedByStudyDay | Frequency counts of prescibed medicines on a particular study day under each disease category. |
| SeasonDisease | Frequency table for individual diseases by disease category, rutu and gender. - This analysis should provide insights into seasonal variations of diseases |
| SeasonMedicine | Frequency table for prescribed medicines by rutu and gender. - This analysis should provide insights into seasonal variations of prescriptions of treatments |
| Box\_AgeMed | Box plot representation of age for each prescribed medicine by gender. - This analysis should provide insights into age groupings |
| Slopegraph\_disPatients | Line chart for each disease by calendar year by gender. - Count of distinct patients is plotted on y-axis, the calendar years are displayed on x-axis. - The x-axis can be expanded to an individual month or week or a day to understand the number of patients at a specific time point. - This provides an easy comparison on similar or dissimilar reporting of a specific disease across gender. |
| DisType\_Diseases | Frequency table for individual diseases by disease category and gender Prameha, Madhumeha and Sthaulya are top 3 most frequently reported diseases. - Prameha and Madhumeha are reported more by males than females. - Vaatavyaadhi - Sandhigata Vaata, Vaatavydahi, Vaatavydahi - Gridhrasee, Sthanabhedena Shoola - Katee Shoola and Sthanabhedena Shoola - Katee Graha are top 5 reported diseases in RMSD. - Vaatavyaadhi - Sandhigata Vaata, Amavaata, Vaatavyaadhi - Asthigata Vaata, Vaatavyaadhi - Vaatakantaka are reported more by females compared to males. |
| Slopegraph\_disVisit | Line chart similar to Slopegraph\_disPatients. This visual shows number of distinct visits to the hospital. |
| MedicineByDay | This visual uses a derived variable for prescribed medicine. The medicines are classified into different kinds, Aristham, Asavams, Bhasmas, Arkas, Dhara, Drops, etc. - Frequency counts by each day is plotted by gender. - Arkas, Avagha, Bhasmas, Panchakarma are prescribed in lesser frequencies. - Kashayam, Aristham, Rasaynama, Abhyanga, etc. are prescribed in higher frequencies. |

## 2. Viz name: Diff\_Visit\_Studyday

| **Sheet name** | **Description** |
| --- | --- |
| Sheet 1 | These boxplots provide the different between 2 visits in terms of days. Patients visit hospital as and when they need to or based on the guidance given by Vaidyas. - There are some patients visiting the hospital for more than 2500 days. - There are some patients only coming for a single visit. - 1 unique date for a patient is considered as 1 visit. - Difference between 2 consecutive visits is calculated for each patient and then plotted as a boxplot. - The observed median difference in 2 visits reduces as the visit number increases. - The median difference starts at 447 days and reduces to 342, 293, 235, 211.5, 197, 191.5, 154, and so on. - Till visit 82 there are 10 or more patients.  This would provide useful insights into patient behaviour w.r.to visiting, recurrence and follow-up with vaidyas, operational challenges. |

## 3. Viz name: Diff\_Visit\_Studyday\_ByDisease

| **Sheet name** | **Description** |
| --- | --- |
| Sheet1 | These boxplots provide similar information to the "Diff\_Visit\_Studyday" visual. - For same patient, find out if a disease is reported more than once on different visits. If yes, then find out the difference between 2 consecutive occurences, use that "duration" in the boxplot. - Display this difference as per visits for each disease reported. - Many diseases are not reported more than once for the same patient.  This analysis should provide a clinical view as well as operational view on follow-up visits for patients. In case of diseases getting cured and a new episode appearing for a patient, large gaps could be seen. If the diseases can be classified into chronic / acute, easily curable vs. difficult to cure then this analysis would provide more useful insights. |

## 4. Viz name: Allopathic\_diag

| **Sheet name** | **Description** |
| --- | --- |
| ICDFreq | The existing ACD codes are used to map to the ICD 10 dictionary. - Frequency table by ICD code, high level term by gender. |
| Baseage\_box | Box plot of baseline age by ICD code and gender |
| All\_vis\_box | Boxplot of total number Visits by ICD code and gender |
| Hospital\_duration | Boxplot of hospital duration by ICD and gender |

## 5. Viz name: Diff\_Visit\_Studyday\_ByMedicine(NotCoded)

| **Sheet name** | **Description** |
| --- | --- |
| Sheet 1 | These boxplots provide similar information to the "Diff\_Visit\_Studyday" visual. - For same patient, find out if a prescribed medicine is reported more than once on different visits. If yes, then find out the difference between 2 consecutive occurences, use that "duration" in the boxplot. - Display this difference as per visits for each prescribed medicine. - Many prescribed medicines are not reported more than once for the same patient.  This analysis should provide a clinical view as well as operational view on follow-up visits for patients. |

## 6. Viz name: 01\_Primary\_madhumeha:

This is an attempt to understand patients with primary diagnosis of madhumeha "M2.0", there are approximately 1400 patients in this subset.

| **Sheet name** | **Description** |
| --- | --- |
| Dis+Med\_Caldate\_view | This is a listing for individual madhumeha patient. - Patient ID, gender, Baseline age are displayed. - Diseases and medicines are displayed one below the other under Code and description columns. - The duration is displayed as per calendar days. - Duration of each disease and prescribed medicine is calculated as difference between maximum date - minimum date + 1 (this may not always be accurate, as it combines 2 independent episodes into 1). - In the tooltip, additional information about total duration of visits to hospital, duration of disease and duration of prescribed medicine are displayed. |
| Dis+Med\_Studyday\_view | This is a listing similar to the calendar view, in place of calendar dates, the study day starting from 1 to maximum study day are used. |
| Box\_MedAge | This boxplot display shows the summary statistics for age of Madhumeha patients getting prescribed to different medicines. |

This individual patient listing would help in understanding relationship between treatments, diseases, co-occurences of diseases and co-admininstritation of treatments.

## 7. Viz name: 01\_\_Primary\_Gridhrasee:

This is an attempt to understand patients with primary diagnosis of Gridhrasee "V2.23", there are approximately 2000 patients in this subset.

| **Sheet name** | **Description** |
| --- | --- |
| Dis+Med\_Studyday\_view | This is a listing for individual Gridhrasee patient. - Patient ID, gender, Baseline age are displayed. - Diseases and medicines are displayed one below the other under Code and description columns. - The duration is displayed as per the study day starting from 1 to maximum study day are used. - Duration of each disease and prescribed medicine is calculated as difference between maximum date - minimum date + 1 (this may not always be accurate, as it combines 2 independent episodes into 1). - In the tooltip, additional information about total duration of visits to hospital, duration of disease and duration of prescribed medicine are displayed. |

## 8. Viz name: 03\_typesOfassessment:

The hospital database captures information for each and every visit. This covers the operational data and clinical data. There are more than 100 Case Report Forms (CRF) within our database with more than 500 variables. Some of the CRF pages are not used at all. These CRFs are covering the In-patient as well as Out-patient visits. This information has been classified into the following categories:

* Ayurvedic data
* Background data
* Diease data
* Doctor's Notes
* Food / Exercise
* Hospital Visit
* Lab report
* Measurement
* Treatment - IP
* Treatment - OP
* Treatment - Procedure
* Treatment / Procedure

| **Sheet name** | **Description** |
| --- | --- |
| TypesOfassessments - Visit | This listing shows data for each patient for each of the categories created above for each visit. - If the data is collected for a particular visit then a vertical bar is presented, if the data is not present then a blank is presented. - Disease data, Treatment data, Treatment - procedure data is captured for almost each and every visit. - Ayurvedic data, Background data, measurements, Doctor's notes are not captured consistently. |
| TypesOfassessments - StudyDay | This listing is similar to earlier listing, it is presented by study day. |
| Summary - Visit | This table displays frequency counts of unique patients for each of the categories created above for each visit. - A patient is counted if the data is non-missing for a specific category. |
| Summary - StudyDay | This table is similar to earlier table, it is presented by study day. |

## 9. Viz name: 080\_medicine\_dis\_all\_met\_rmsd\_prop:

Dashboard 1 in this visual is explained below.  
080\_medicine\_dis\_repeat\_prop\_cumulative:

| **Sheet name** | **Description** |
| --- | --- |
| Dashboard 1 | This shows individual patient data for disease and treatment for Metabolic and RMSD patients. It provides diseases and treatments per patient as either "disease reported 1st time" or "repeat", "treatment reported 1st time" or "repeat". - It is reported by studyday (or visit) when a disease and medicine is reported in the data. |

This provides the following information:  
(1) When a new disease is reported, usually a new treatment(s) is (are) reported  
(2) If there is only a new treatment added then it could indicate, the earlier treatment may not have worked, or it explains the nature of treatment regimen.

## 10. Viz name: 080\_medicine\_dis\_repeat\_prop\_cumulative:

Dashboard 1 in this visual is explained below.

| **Sheet name** | **Description** |
| --- | --- |
| Dashboard 1 | This dashboard should be read in parallel to the Dashborad on **080\_medicine\_dis\_all\_met\_rmsd\_prop**visual. - Individual patient data is presented by visit in a cumulative manner for disease and prescribed medicine. In the adjoining table % are displayed. |

## 11. Viz name: 080\_medicine\_dis\_repeat\_prop:

Dashboard 1 in this visual is explained below.

| **Sheet name** | **Description** |
| --- | --- |
| Dashboard 1 | This shows individual patient data for disease and treatment for Metabolic and RMSD patients. - The top section provides diseases and treatments per patient as either "disease reported 1st time" or "repeat", "treatment reported 1st time" or "repeat".  - The section below displays detailed data for diseases and prescribed medicines. |

## 12. Viz name: 085\_dis\_1st\_time\_refCal\_NodesEdges:

Each of the 106 diseases (10 Metabolic and 96 RMSD) is considered as a reference disease.

* Day 1 is calculated as the reference day 1 for individual patient for each disease.
* Other diseases for the same patient are positioned either before or after compared to this reference disease.
* Duration w.r.to this reference day is calculated before and after day 1. This calculation provides the background view as well as future view.
* This referencing allows for more informative background disease as well as background medicine information.
* The duration is split into the following time points:

| **Before** | **After** |
| --- | --- |
| Day 1 as reference |  |
| Before 1 month | Within 1 month |
| Before 2 months | Within 2 months |
| Before 3 to 6 months | Within 3 to 6 months |
| Before 7 to 12 months | Within 7 to 12 months |
| Before 2nd year | Within 2nd year |
| Before 3rd year | Within 3rd year |
| Before 4th year | Within 4th year |
| Before 5 year | Within 5 year |

* 1 sheet for each reference disease is created.
* Frequency count of diseases and prescribed medicines is displayed.
* Prior counts are displayed in red colour and After counts are displayed in Green colour.

This view should provide good insights into the causal relationships.

## 13. Viz name: 085\_dis\_count\_edges\_3rd\_byPeriod02try:

Dashboard PrimaryDis\_relatedDisMed (2): circular view of disease and medicine relationship.  
This view allows the following comparisons:

* Relationship between diseases and treatments across different time points.
* If a disease is experienced in different time windows then would the treatment options look different or would they look similar.
* Occurrence of diseases and proximity -- do the diseases precede and / or succeed each other, etc.

E.g. Amavaata, what all diseases were experienced and treatments given before the 1st occurrence of disease and after the 1st occurrence of disease. How far or how close were these events are given in terms of within 1 month, within 2 months, within 3 to 6 months, 1 year, 2year etc. on both sides.

* The inner circle displays the diseases.
* The outer circle displays the treatments.
* Counts of distinct medicines prescribed and distinct diseases experienced are given.
* More frequenct counts are displayed in the table above for each of the periods.