Talking points

|  |  |
| --- | --- |
| Slide | Points |
| 1 | Thank you all for attending today’s session, I would to take you through my PhD thesis titled as “Analysis of hospital based ayurvedic clinical practice to gain real world data knowledge”  Ashwinis and Girish were the guiding force behind the work carried out over the years. |
| 2 | Agenda slide covering important sections of presentation |
| 3 | A brief background on the setting up of this study within TDU and IAIM hospital.  Thank you to both Poornima and Prasan for their role from Hospital management team. This study was a retrospective analysis of the available data.  Moving onto the next slide now, let us go to the main part of today’s presentation focusing on the background and previous research, methods employed in the study, results seen via various analyses, and finally conclusions drawn at the end of this experiment. |
| 4 | Starting with the background and previous supporting research section: |
| 5 | Various electronic equipment like computers, mobile devices, wearables, and other sensors collect and store huge amounts of health-related data. This explosion of data carries potential to better design and conduct clinical studies to answer questions previously thought infeasible. Advancement of cutting-edge analytical capabilities is allowing researchers to analyze and comprehend this data at greater depths, permitting medical product development and approval at an accelerated speed [1].  International Data Corporation (IDC) is one of the premier global providers of market intelligence, information technology, and a host of other areas. They predicted in a report issued in Dec 2018 that the world’s cumulative data will grow from 33 zettabytes to a 175ZB by 2025, for a compounded annual growth rate of 61%. A zettabyte is a trillion gigabytes multiplied that by 175 times. This growth of data has been seen in every industry, in every corner of the world.  Real world data (RWD) is the information relating to patient health status and/or the delivery of health care routinely collected from a variety of sources like epidemiological studies, clinical practice, already published articles to answer questions previously thought infeasible.  A study carried out by Clarivate Analytics, USA, reports 27 (non-exhaustive list), <5% of all approved drugs, examples of drug approvals by US FDA, EMA, Japan’s PMDA and Health Canada, across broad spectrum of medicines between years 1998 and 2019 using RWD from Electronic Health Records and registries. |
| 6 | Approval of Ibrance by US FDA for male breast cancer, is one such recent example of approvals using RWD data. The company did not have carry out a specific study on male breast cancer patients, but RWE data was relied upon.  These data were used either as primary data, when non-comparative data were available to demonstrate tolerability and efficacy, or as a supportive data when validating findings. |
| 7 | EMA one of the leading health authorities in the world has published a view on RWE and its approach towards big data in a paper called as “Strategic reflection regulatory science to 2025”.  A drug already approved for females and French health authorities allowing a Real World Evidence (RWE) study of 600+ patients, over a period of 18 months, for a conditional re-imbursement scheme in COPD.  This provides increasing usage of “naturally reported data” in drug approvals in modern biomedicine. These examples provide evidence of novel use of data, which may have otherwise gone unused. The power available to society would have never been unearthed if not for this way of use of RWD [2]. |
| 8 | After looking at some examples from the modern medicinal world, let us focus on potential areas in Ayurveda:  Is Ayurvedic area dealing with the same type of challenge of not realizing the potential of available data? Just to give a glimpse of enormity of data: data for more than 10 crore number of patients have been reported on AYUSH website (As of May 2020). More than 140+ countries have population of less than 10 crores [3].  It is safe to assume that the conceptual developments in Ayurvedic knowledge base have taken place through everyday observations and basic laws of nature. These fundamentals have been adjusted to the relevant times as per the passage of time based on observations and experiences, where there are no artificial restrictions on usage of medicines, duration of treatment or type of patients to treat, which is next to impossible in a protocol driven clinical trial setting [4] [5].  Would it be worthwhile and would it be possible to carry out a Real world evidence study within ayurveda? We have attempted to do so – the subsequent sections explain the methodology used for the same. |
| 9 | Let us focus on the methods section now. |
| 10 | Study Aims and Objectives are as follows:  Aims: To generate real world evidence in the field of ayurveda by using tools and techniques like visual analytics and deep learning methods  Objectives:   1. Converting clinical life data into analyzable format 2. Clinical data understanding 3. Studying demographics and patient specific factors 4. Diagnostics and treatment data 5. Outcome and effect   The next slide explains the practical steps used in getting to the aims and objectives. |
| 11 | Taking inspiration from respected Prof Patwardhan’s quote, “Charaka would not have ignored modern technologies if they had been available during his time” [6], this study attempts to discover hidden wealth of Ayurveda related information in EHRs created at TDU hospital using modern methods of data sciences and statistical programming. Since 2011 to October 2017, the hospital database contained data for approximately 50,000 patients, more than 1,70,000 visits, close to 900 disease types and more than 3,000 variations of medical procedures [7]. The study was approved by the authorities of IAIM and TDU. This study targets the methodological and learning framework as well as creation of many tools based on free technologies for various stakeholders in following categories:   * Hospital managements, clinicians, and patients * Universities and learning institutes – clinical communication, researchers to build vital evidence-base * Policy makers – AYUSH and relevant ministries * Healthcare providers - Ayurveda Healthcare systems, General healthcare systems |
| 12 | As we are dealing with patient level data, we took care of the data flow steps in the following manner –   * An associate from Medical Record Department (MRD) helped in extracting the data in anonymized format. * The existing database contains sensitive information related to individual patients. All the efforts were made to protect sensitive information like   + Name of patient,   + Full address,   + Phone numbers,   + Socio economic status, etc. * This information was not extracted from source data in order to maintain confidentiality.   The hospital database is a live system and as the data is updated on a daily basis, we had to make a decision and carry out analysis on a particular version of data. We have used the data from year 2011 to 2017. On relevant slides, this version of data is reflected for context. |
| 13 | This slide depicts the pictorial representation of the data flow from source to analyzable data to converting the data into reports of different kinds.  The source data from operational system, clinical system, and coding dictionaries is combined logically. Along with the source information, there are a few additional files created using subject matter expertise. These are logically added to the other source datasets. These give rise to intermediate datasets created for calculations and transformations. Some of these are temporary datasets and some of these are stored in a staging area used by analysts. Staging area is an area dedicated to individual project. Final output datasets are stored in a central place called as a data warehouse. It is a central repository of data within for an organization. These datasets are used to make operational, business, and scientific decisions by various stakeholders by converting them into interactive analysis, dashboards, formal project reports. |
| 14 | Data was analyzed through SQL [90] and R programs [91], python [92], Java [93], D3js [94] and tableau [95] software. Tableau is used to generate final visualizations.  These various analyses enable data to be reported in different levels of details. Most of these representations are interactive. End users can perform filtering tasks while using the visualizations. Tableau’s drill down facility provides additional ways of analyzing the data. Tooltip functionality allows extra dimension to provide more details [95].  This is more as “for your reference” – as how many different technologies were used along with the underlying “subject matter – ayurvedic medical data” |
| 15 | Let us not spend a lot of time on this slide – but this slide is kept here to depict how data was captured by the TDU hospital and complexity attached to saving data in various structured tables. There are ~150 different data tables developed as a “system architecture” to hold hospital management information system data.  The data tables listed in yellow color are used for the analysis and are combined logically to generate coherent version of patient disease and medical journey. |
| 16, 17, 18 | In next 3 slides, we highlight technical details about the hospital database: Out of more than 150 tables approximately 20 to 25 tables were used to generate datasets useful for further analyses. Subsequently we displayed flowcharts outlining ~30+ steps to go from Live source to Staging data to transformed data to ~30+ source variables + ~30+ derived variables in 01adsl\_met\_rmsd dataset: patient level data covering treatment and disease information.  The Orange part of the table in slide 18 covers the additional variables derived for various analyses.  If we are to repeat this type of analysis for some other hospital, similar workflow would be needed to combine data from different sources. The learnings gained from this exercise will be very helpful.  Technical work so far has formed the basis of possible operational and clinical analyses going forward. |
| 19 | Moving forward to slide 19, we outline various methods employed in the study on this slide – none of these methods is possible w/o the electronically captured data in the source database and technical work carried out so far to combine data.  Sections so far cover the journey from an individual observation to a logically developed dataset capable of handling various kinds of analyses going forward.  This covers the methods section at a high level. |
| 20 | Before formally starting work on this thesis - preliminary analysis was carried out by Dr. Girish Tillu. The table showed on this slide is part of that analysis. It is showing a lot of patients in Metabolic area and (Rheumatic and Musculoskeletal disease) RMSD area, hence major analysis will be carried out using these 2 disease areas. |
| 21 | Based on the ACD classification, which was developed at C-DAC, we created “disease grouping”. There are 10 metabolic diseases and 97 RMSDs.  Similar methodology could work on any other disease area as well, some work has been carried out Parkinson disease, Chronic Kidney Disease (CKD), details are not included at this point in this slide deck. |
| 22 | Let us now focus on what type of results did we found, we have split the results in a few broad categories – (1) Clinical data understanding, (2) Demographics and Patient specific factors, (3) Diagnostics and Interventions  The thesis document covers close to 50 different analysis ideas, due to time constraints, in this presentation I would cover 10 to 12 examples of analyses reported in the thesis – let us go through them to cover the results section. |
| 23 | Subsequent section shows examples of analysis related to clinical data understanding. |
| 24 | Patient analysis by country – a Country-wise visualization on the world map  Interpretation:   1. This graphic shows the geographical distribution of patients. 2. In this version of the data, there are patients coming from 50+ different countries. 90% or more patients are from India and remaining 10% patients are from different parts of the world. 3. If we are to focus on India specific analysis, if the individual state and city information is available then additional drill down representation is possible – this picture will allow us to identify the distribution of patients and diseases from different parts of India. 4. A very good tool for understanding geographical trends   Use for each stakeholders:  A pictorial representation of the world data, very useful as an executive summary. This form of data representation will help any public health official. More details related to diseases, treatments, additional demographic characteristics could be added to the tooltip to efficiently recover key information as and when needed. |
| 25 | This slide shows a tabular summary of Blood group distribution by gender:  Interpretation and use for stakeholder:   1. Blood group distribution for such a large number of patients is a great source of knowledge. Even though this does not help in every day treatment options, there is undoubted epidemiological value in this tabulation. 2. There are obvious mistakes in documenting the blood groups observed via this tabulation – showing another secondary use of this tabulation to build data quality related efficiencies. |
| 26 | Visit pattern analysis: Frequency counts of 4 parameters, (1) new Out-Patients added on that day, (2) total number of patients visiting on that day, (3) total number of In-Patient visits on that day, and (4) total number of Out-Patient visits on that day were calculated for each day to understand the patient flow to hospital from year 2011 to 2016. The calculated information was represented on a calendar.  Data / observations: In a competitive world, understanding how patients plan their visits to a hospital and their social demographics can help in running an operation which would make the patient as the focus point of efficient operations. Operational efficiency allows businesses to build an edge and it is applicable in any field. Visit pattern analysis (Figure *3‑3*) allows insights into the same.  Even though the current slide does not show data across years, this visualization depicts data from the very first day of the hospital in 2010 on a calendar. In Feb 2011 there were more than 50 patients visit on a single calendar day. In May of 2011, there were more than 200 patients visit on a single day. Otherwise, the average number of visits was hovering around 30 – 50 patient visits.  The underlying data shows that the patient inflow increasing year on year. Out of 2100+ days of data from year 2010 to 2016, there were 158 days on which more than 100 patient visit days were there, which accounts for 7.33% of all the days. Most of these 100+ patient visit days happened on Saturday, showing patients’ tendency to visit hospital on a weekend. |
| 27 | Next section shows examples of analysis related to Demographics and Patient specific factors. |
| 28 | In-depth review of number of diseases. Following interactive report was created and was analyzed for the complete set of patients, not subsetted for metabolic and RMSD groups.  Interpretation:   1. Almost 12k+ and 14k+ patients with only a single disease, these patients could have come only once to the hospital and may not have come back at all after reporting the 1st disease. Are these patients largely coming in for “2nd opinion”? Or if this data is to be looked at positively, they are getting benefitted and hence are not coming back for consultation beyond the first reported disease? 2. The summary statistics for each of the category across number of diseases looks quite similar. 3. Patients having more than 10 disease conditions. These patients could be having a lot of faith in Ayurvedic treatment, for them to continue on, they could have found the underlying treatment effective. The maximum age of the 108 is a possible data issue. While finding data issue was not a primary outcome of the analysis, there is this secondary usage available to the scientific community. Similar anomalies are seen in a few other groups, e.g., patients reporting 23 diseases, is this accurate or needs additional data checks?   Use for the stakeholders:  Is there a similar tabulation available for another Ayurvedic hospital, or any other private or public hospital? Is there a similar distribution viewed? This can be used to understand the use and misuse of the limited medical sources across the geographies.  Data / observations: Data quality is an inherent requirement for any data related exercise. Some insights based on patterns of missing data, outliers and potential scientific and medical inconsistencies can suggest methods to improve the data quality which would be of essence to hospital administration.  Analysis covering descriptive summary statistics by number of Diseases by age and gender was carried out (Figure *3‑11*) and explained in detail in the earlier chapter 3 section 3.3. This table summarizes descriptive statistics for age. The first column categorizes how many diseases were reported per patient going from 1, 2, ..., n diseases. The table shows 23 diseases as the maximum number of diseases reported by a patient. There is a large part of data showing “NA” = unknown number of diseases. Both observations point to potential data issues. |
| 29 | This analysis shows the duration of hospital visits:  A tabular representation of Patient visit duration for Disease categories by Gender (Figure *3*‑*14*), using the following logic: The duration between the first visit and the last visit for each patient has been calculated and categorized as follows: >= 1 day, >= 1 month, >= 2 months, >= 3 months, >= 6 months, >= 1 year, >=2 years, >= 3 years, >= 4 years and >= 5 years. In this analysis patients were counted multiple times as per available data for each time period. E.g. a patient visiting for more than 5 years was counted in all categories. If a patient discontinued in the 4th month then that patient was counted in Day 1, >=1 month, >=2 months, >=3 months categories. The colour gradient moves from Red to Green denoting low to high number of patients in each category.  Columns show: 1 = Patients with at least 1 metabolic diseases, 2 = Patients with at least 1 Rheumatic, Musculoskeletal (RMSD) diseases, 99 = Patients with at least one disease from each of the groups  Interpretation:   1. This summarizes the duration of the patient and hospital association for 14000+ patients 2. Patients from RMSD disease are more prevalent and ~15% patients visit the hospital after 1 year 3. ~ 38% patients come to hospital for more than 1 month. 4. Smaller duration of patient and hospital association may mean either the patients are benefitted by the treatment or are not happy and hence discontinue the treatment. 5. Longer duration of association may mean that the patient is receiving benefit and hence is coming for regular follow-ups for the same condition or the disease condition could be chronic in nature. 6. More patients having reported at least one disease from each of the groups seem to continue for longer duration as compared to patients having either RMSD or metabolic diseases.   Observations are similar to what we saw in the earlier descriptive statistical analysis. |
| 30 | We all know that there are disease variations experienced due to changes in rutus. This table shows frequency count of patients across rutus. Some of the chronic diseases within metabolic area may not be showing any signs of upward or downward trends. But some of the RMSD diseases must be reported in either higher or lower frequency.  This analysis would be useful from various angles: which medicines to keep in the pharmacy, which specialist vaidyas to be employed in which season – of at least have a visiting vaidya in appropriate season. Have disease specific camps, to name a few.  Presentation of disease burden by gender, Indian seasons (rutus) and disease category provides data about possible variations reported for different diseases (Figure *3*‑*15*). (1) Prameha, (2) Madhumeha, and (3) Sthaulya were the top three most frequently reported metabolic diseases where as (1) Vaatavyaadhi – Sandhigata Vaata, (2) Vaatavyaadhi, (3) Vaatavyaadhi – Gridhrasee, (4) Sthaanabhedana Shoola – Katee Shoola and (5) Sthaanabhedana Graha – Katee Graha were the top five most frequently reported RMSD diseases. Prameha and Madhumeha were reported more by males than females. There were more female patients with disease condition Sthaulya. In general, RMSD diseases were reported in more females than males. For RMSD disease group, 51 out of 97 diseases were reported in <= 10 patients. Metabolic diseases were not varying across seasons, while RMSD diseases had some seasonal variations (Figure *3*‑*15*). |
| 31 | Last section within results shows examples of analysis related to Diagnostics and Interventions. |
| 32 | Following 3 slides provide a view on patient journey or longitudinal view. Two longitudinal interactive views were created to display individual patient data.  This slide shows: first version of patient profile contains the following information (Figure *3*‑*4*): Patient ID (mr\_no), gender, study day, In-Patient visits are displayed in blue colour and Out-Patient visits are displayed in Orange colour. The tooltip of the interactive display holds information about the following data points not displayed on the page: (1) Study day, (2) Total duration of hospital visits, (3) Disease description variable accompanying ACD codes, (4) Medicine provided at that visit, (5) Minday Metabolic: First day on which any metabolic disease has been reported by patient, (6) Minday RMSD: First day on which any RMSD disease has been reported by patient.  Multiple representations of data allow the end user to review data with different perspectives. Patient profile reports provide detailed view of individual patient’s disease condition, prescribed medication, co-morbidities along with basic demographic information. Treating doctors and researchers will greatly benefit from this visual display. This representation (Figure *3*‑*4*) provides the patient an understanding of the disease chronology as well as the prescribed medication and progress. Interpretations drawn from these representations and potential approaches to improve current version of patient profiles used at the hospital are provided in chapter 4 sections 4.2.8.  Data / observations: If we can understand the patient journey which includes, diagnoses, treatments, outcomes over time, it will allow the treating doctor to adjust and do better for the patient. The insights can as well give more understanding of the disease and the science behind it. This in-depth analysis shows us how to combine three analyses to understand individual patient journey. Figure *3*‑*5*: Patient visit profile – Vertical view, Figure *3*‑*27*: Patient Disease and Treatment administration by Study Day, and Figure *3*‑*29*: Patient Cumulative Disease and Treatment administration by Visit are combined to show how can a treating doctor use these three tools simultaneously. |
| 33 | Second version of patient profile contains the following information (Figure *3*‑*5*): Patient ID (mr\_no), gender, base age, category, Code, description, study day. The Diseases were displayed in blue coloured bars and treatments prescribed were marked in orange coloured bars. Disease duration and treatment duration bars were created as follows: Duration between minimum and maximum reported date for a disease as well as prescribed treatment was calculated, this duration was displayed on the visualization. The tooltip contains information about the following data points not displayed on the page: (1) Daystt: Start of event in days, (2) Disdur: Duration of event in days, (3) Disstt: Start date of event, (4) Diend: End date of event.  Subset these 3 analyses for patient MR000774. This data is for a 68-year-old female patient who has been coming to the hospital for close to 3 years (960 days). In this period, she has visited hospital for 29 times. There were 9 different diseases reported and 35 prescribed medicines in the hospital database. For a few of these diseases ACD code has not been reported, giving rise to missing data. Madhumeha and Prameha have been reported as first two diseases reported. Is this a correct representation having both Pramhea and Madhumeha being reported?  Diabetes condition is treated by DNil and Nisakathakadi kashayam. Prameha was treated by Nisakathakadi, kathaka kathiradi. Nisakathakadi was prescribed as Choornam / powder and Kashayam, Nisakathakadi treatment was stopped and then Glokustat and Diabecon DS were prescribed. This line-by-line data review provides information about treatment protocol being followed. |
| 34 | The analysis carried out in Figure *3*‑*29* helps in understanding first time disease reported and repeat disease representation may be used to understand the co-morbidities. 26 times diseases were reported and only 9 were unique showing a smaller number of complications, if we ignore “Not coded terms” then there are only 5 unique reported. 95 times treatments were prescribed out of which 35 were unique. There are 2 vaatavyadhis, arthritic conditions, reported after a few visits. Similar details about the treatment can be found in the medicines section.  The analysis carried out in Figure *3*‑*27* provides a tabular and line-by-line listing view at one go, providing one more variation of the same analysis. This report shows reported diseases and medicines as well as updates across timepoints. These three patient journey analyses should be good additions to existing tools to study patient status before a patient visit for any doctor.  Insights: The spelling was different in different prescriptions which does not change the underlying treatment but creates a perception about extra treatments prescribed to tackle the same disease. If this can be controlled by the coded medicine dictionary and type of formulation, then this will help in data analysis. The arthritic conditions reported by the patient, could be just age-related ailments reported by the patient. This elderly female patient does not have many co-morbidities reported which is a good sign from health point of view. |
| 35 | This slide shows a co-morbidity analysis.  Co-morbidity analysis was carried out by using 3 different approaches, on this slide, we will understand the first approach: First approach produced a bubble plot, boxplots, summary statistics tables and frequency tables for number of other diseases reported along with the primary disease.  Algorithm for first approach: (1) A unique combination of Patient ID, gender and reported disease at any given time point was created. (2) Subsequently, a dataset having combination of diseases for an individual patient was created. E.g., if a patient had reported 5 unique diseases, then all the combinations of these 5 diseases were created i.e. 5C2 combinations were created i.e. 10 combinations. (3) The resulting data had the following structure: Patient ID, Disease1, Disease2, and Gender. (4) Frequency count of distinct patients was calculated for each Disease1, Disease2 combination and gender. (5) Using this data following analysis was carried out:  Summary statistics of age group for each disease by gender; Boxplot of age group for each disease by gender; Bubble plot for each disease where the bubble size was determined by the count of unique patient IDs. For each disease number of other unique diseases by gender were reported. Tooltip on the bubble plot provides information about count of distinct number of patients, and summary statistics for age group.  The dashboard is controlled by a “Primary Code” or a reference disease and relevant data is displayed on the page. Other bubbles in the bubble plot, display the diseases reported by this subset of patients at any point in time (these could be clinically related or unrelated or could have occurred before or after the occurrence of reference disease). The tooltip shows minimum, median, and maximum age and distinct counts of patients. A table on the left side shows number of other diseases experienced by the patient (Figure *3*‑*22*, Figure *3*‑*23*, Figure 3‑24).  This display provided a comprehensive view of the disease clusters. Comorbidities were easily identified, some of them are clinically relevant, and some of them are not. Bubble size provided comparative view of number of patients reporting a specific disease. The age group distribution for each gender was available. Some diseases were reported more by males or by females, easy to spot on the graph. Box named “Number of other diseases” provided a contextual display about number of co-morbidities. Some diseases had higher number of co-morbidities, some had lower number. Variations were seen amongst gender as well. This analysis did not consider the before or after nature of time points, hence did not provide insights into the causal relationships between diseases.  Some examples like prameha (causing many diseases for both the genders), pandu roga (mainly reported by females with many disease, and relatively low numbers reported by males), sandhigata vata (reported by more females), etc. have shown that meaning of shlokas can be shown in the modern data format. This type of exercises can be carried out with help of ayurvedic experts (Figure *3*‑*22*, Figure *3*‑*23*, Figure 3‑24, Figure *3*‑*25*, and, Figure *3*‑*26*). |
| 36 | The underlying data was generated from every day medical practice at the hospital. Hence the diseases were reported almost at random. The following analysis used first occurrence of any disease as day 1 for an individual patient. Using this as a reference day “before period” and “after period” was derived. “Before period” provides significant amount of “baseline data”, “after period” provides specific insights into what would happen after the onset of the reference disease. The following algorithm was used to create the underlying data for analysis:   1. Each of the 107 diseases (10 Metabolic and 97 RMSD) was considered as a reference disease. 2. Day 1 was calculated as the reference day 1 for individual patient for each disease. 3. Other diseases for the same patient were arranged either before or after compared to this reference disease. 4. Duration was calculated before and after day 1, which is the reference day. This calculation provided the background view as well as future view. 5. This referencing allowed for more informative background disease as well as background medicine information. The duration was split into the following time points as described on the slide |
| 37 | Existing circular analysis:  Analysis for disease – treatment with pre and post visit window approaches:  The circular visualization allows a single page view of relation between disease – disease and / or disease – treatment across multiple time points. This view shows the following information: (1) A table on the middle row: On day 1 of a disease how many distinct diseases have been reported and how many distinct medicines prescribed, this same information is shown as the green bars inside a circle, (2) Pre and post time windows are displayed and for each of the time window a similar table is represented in the upper section of the visualization. (3) In the lower section of the visualization, 1st row represents the co-occurrence of disease – disease and / or disease – treatment before day 1 of the reference disease. (4) Last row represents the same co-occurrence data after day 1 of the reference disease (Figure *3*‑*37*, Figure *3*‑*38*).  Visualization of  Co-occurrences of disease and medicine  Co-occurrences of disease and disease  Interpretation from additional Disease – treatment analysis with pre and post visit window approach are as follows: in circular data representation many green lines means that there is a greater chance of diseases reported by patients, there is a greater chance of a medicine prescribed for a disease. If there are very few lines then the combination is clinically not meaningful or if it is meaningful then it is a very rare combination which needs to be studied further (Figure *3*‑*37*, Figure *3*‑*38*). On a single page there are multiple dimensions of the disease – disease and / or disease – treatment combinations are shown. |
| 38, 39 | An attempt was made to understand the disease trajectories for patients by using mathematical distances. There are numerous distance measures available in mathematics and statistics which allows understanding of similarity and dis-similarity between objects, in our case disease trajectories [110].  Following assumptions were used to derive the disease trajectory: (1) Diseases experienced by each patient were sorted by date and only first instance of a disease was retained. (2) This enabled in creation of a disease trajectory for each patient for each reference disease, before and after the occurrence of the reference disease. (3) Cartesian product of patients was created for each reference disease, so that distances could be calculated. A cartesian product is a set of all possible pairs, in this case all possible pairs of reference disease and other diseases for individual patient. (4) The similarity measure was calculated for each disease trajectory, e.g., Jaccard distance was used as a distance measure for this display [111]. The Jaccard distance highlights similarity between finite sample sets. It is defined as the size of the intersection divided by the size of the union of the sample sets [112]. For our example, this will be calculated as the common diseases reported divided all the diseases reported in each case. (5) Jaccard distance closer to 0 shows dissimilarities and closer to 1 show similarities. (6) The distances were divided into 4 categories 0 to 0.25, 0.25 to 0.5, 0.5 to 0.75 and 0.75 to 1 for data visualization perspective. (7) These calculated distances are displayed as a butterfly plot for easy comparison of underlying values [105]. This plot is a comparative bar plot to display comparison of a continuous variable across groups. Similar Analysis to understand the medicinal trajectory was performed.  Distance score-based analysis reveals the following observations: More number of patients with Jaccard distance closer to 1 was seen for the Post reference day 1 period. This could be pointing to similar biological activity caused by a particular disease. This could be a very important finding from this analysis (Figure *3*‑*39*). In the medicinal display the similarity scores are lower as compared to that for the disease trajectories. Which implies that most of the prescribed treatments are dis-similar for both the periods. It is observed that around 50% of treatments could form the base of treatment regimen and could be same for the patients. The remaining part of the treatment regimen is driven by individual patient characteristics. The before and after medicine trajectories would show such underlying data. E.g., for M2.0, there are very few patients having distance above 0.5 for both genders (Figure *3*‑*40*). This analysis should be executed using other mathematical distances to understand the consistency of results. If the disease classification and treatment tagging in the underlying data is improved then we should be able to see much better results, with lesser confounding effect. |
| 40 | Radar plot representation: a multidimensional, comparative view of the different diseases was created to understand at various aspects of the diseases. The radar plot chart presents multidimensional metrics. Radar plots can convey a large amount of information [100]. They provide a standardized view of different indicators on one scale. The following information for each disease was visualized as a percentile and is represented as a dimension on a heptagon (as there are 7 parameters considered in this example): (1) Distinct number of patients for each disease, (2) Number of times a disease is reported, (3) Number for a specific disease (chronological number of disease reported by a patient) e.g. a disease is reported as the very first disease or third disease or fifth disease, etc., (4) Number of diseases before the specific disease, (5) Number of diseases after the specific disease, (6) Number of treatments before the specific disease, (7) Number of treatments after the specific disease. Trellis plot display allows multiple representations of same kind next to each other [99].  The radar plot shows multi-dimensional data in a short space, 7 different parameters were shown on 7 vertices. Different shapes suggest that there were underlying differences to the data structure.  Radar plot for multiple diseases is shown next to each other. This is showing massive amounts of information immediately. Differing shapes provide differences reported in the data and an easy way of identifying differences. If there is additional data made available in a structured format, then these parameters could also be added on the radar plot. This radar + trellis combination provides a more powerful tool to visualize large amounts of data on a single page (Figure 3‑41).  A heptagon visualizes a disease  Differing shapes provide differences reported in the data  This slide concludes explanation about various analyses performed using different methods. |
| 41 | Let me take you to the concluding section of this presentation |
| 42 | By showing various analyses carried out so far, we are adding Real world clinical practice data analysis as an additional part of the overall evidence base for any medicinal product.    Ayurvedic vaidyas usually use paper-based case report to record a patient’s Ayurvedic parameters along with other details of medical consultation. These are typically not exchanged with other vaidyas. There is a huge amount of data available on paper and if digitized could be a big revolutionary step. Increased use and interoperability with electronic medical records of digital Ayurvedic patient management systems are required. Based on a report published by AYUSH [66], there are 4.5 lakh registered Ayurvedic practitioners. Even if 5% of doctors start using EMRs, i.e., 22,500 doctors and if data for 2 new patients is entered every day (~225 working days) for the whole year, 50 lakh unique patients’ data can be generated in a single year. Currently, this gold mine of data has not been built yet. |
| 43 | Due to the above-mentioned outcomes, the following contributions can be possible:   1. Contribution to Public health data creation based on large data at our disposal which is not marred by artificial boundaries imposed on patient disease conditions and treatments prescribed as followed in a designed randomized clinical trial. 2. Make recommendations to the practitioners for standardized way of data collection, analyses and reporting which will support future EMR based RWD studies 3. Understand the hidden wealth of data for Transdisciplinary expansion of thoughts    1. Sustainable treatment solutions for diseases readily available    2. Thought provoking work to generate new needs through unconventional use of the data    3. Expand the use of modern IT solutions like IT infrastructure, electronic health records, cloud, etc. within Ayurvedic area where appropriate – Ayur IT solutions.    4. Take advantage of freely available cutting-edge software(s) to create new approaches    5. Introduce statistical programming (Ayurdata analyst) as a tool to Ayurvedic area |
| 44 | Slides 44 and 45 cover the potential type of questions could be answered and who could be benefitted? |
| 45 | As I had spoken at the start of the presentation, the study targets the methodological and learning framework as well as creation of many tools based on free technologies for various stakeholders in following categories:   * Hospital managements, clinicians, and patients * Universities and learning institutes – clinical communication, researchers to build vital evidence-base * Policy makers – AYUSH and relevant ministries * Healthcare providers - Ayurveda Healthcare systems, General healthcare systems |
| 46 | This thesis outlines many tools which can be used by various stakeholders. They are free and easy to use. They allow multi-dimensional display of complex data in a very short amount of space. The tools can create evidence for multiple stakeholders. Free softwares like, R, python, Java, tableau and many more have made it possible to harness the power of data in many ways. There is a need to have a profession of a “Statistical programmer” or a “clinical programmer” or an “Ayurdata expert”. This role can contribute to database development, data collection, data cleaning aspects, creating analyses ready datasets, and to finally analyses and reporting. This role should have capabilities related to information technology, data management techniques for generating quality data, in addition to knowing basic and advanced statistical and data science concepts. The computational advances in the world of computer science could be leveraged via appropriate software. Theoretical ideas can be converted into practical interactive visualizations and interactive analyses using multiple technologies. These will help convert individual data observations into summaries then into stories thus enabling knowledge generation.  Ayurdata expert can contribute to creating documents for medical journalism, medical education, medical marketing of healthcare products, publications, research documents, and regulatory documents by collaborating with other experts (Table 5‑1). We believe that this would be a pioneering effort within ayurvedic EMR area. |
| 47 | Potential benefits of this work: The database will be useful in publishing case studies, case series, etc. in a short period of time. This will help us gain more visibility in scientific world:   * The database will become more searchable * More empirical data will be available at our disposal * The data will be closer to analysis ready format * The database could become a model database for other Ayurvedic institutions to follow |
| 48 | Thus contributing to convert data to knowledge to wisdom |
| 49 | This last slide lists down ongoing and future work – asking for help from the community to take this work forward |