Analysis of hospital based Ayurvedic clinical practice to gain Real World data knowledge

Vinay Mahajan

Synopsis

Guides: Dr. Ashwini Godbole,

Dr. Girish Tillu,

Dr. Ashwini Mathur

1 Table of Contents

List of tables

List of figures

Acronyms

ACD: Ayurvedic Classification Dictionary
A-HMIS: Ayurveda Hospital Management Information System
AYUSH: Ayurveda, Yoga and Naturopathy, Unani, Siddha and Homeopathy
CCRAS: Central Council for Research in Ayurvedic Sciences
CDISC: Clinical Data Interchange Standards Consortium
COPD: Chronic Obstructive Pulmonary Disease
COSTART: Coding Symbols for Thesaurus of Adverse Reaction Terms
CTC: Clinical Toxicity Grade
DARWIN EU: Data Analysis and Real-World Interrogation Network
EHR: Electronic Health Record
EMA: European Medicinal Agency
EMR: Electronic Medical Record
EU: European Union
GCP: Good Clinical Practice
GCRP: Good Clinical Research Practice
GDP: Gross Domestic Product
HM: Hospital Management
I-AIM: Institute of Ayurveda and Integrative medicines
ICD: International Classification of Diseases
ICH: International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use
IDC: International Data Corporation
IDC: International Data Corporation
IP visit: In-Patient visit
ISO: International Organization for Standardization
LOINC: Logical Observation Identifiers Names and Codes
MedDRA: Medical Dictionary for Regulatory Activities
NCI: National Cancer Institute, USA
NEJM: New England Journal of Medicine
OP visit: Out-Patient visit
PMDA: Pharmaceuticals and Medical Devices Agency (Health Authority Japan)
RCT: Randomized Controlled Trial
RMSD: Rheumatic and Musculoskeletal disease
RWD: Real World Data
RWE: Real World Evidence
SQL: Structured Query Language
TDU: The University of Trans-Disciplinary Health Sciences & Technology
THERAN: THE Research Application Nexus
TKDL: Traditional Knowledge Digital Library
UHC: Universal Health Coverage
US FDA: US Food and Drug Administration

WHO: World Health Organization
WHO-ART: World Health Organization Adverse Reactions Terminology
WHO-DDE: World Health Organization Drug Dictionary Enhanced

Preface

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]. 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.

Approval of Ibrance by US FDA for male breast cancer, 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, are a couple of recent examples of approvals using RWD data. 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. 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. 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].

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: 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].

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 51,000 patients, more than 1,50,000 visits, close to 900 disease types and more than 3,000 variations of medical procedures [7]. The proposed study "Analysis of hospital based Ayurvedic clinical practice to gain Real World data knowledge” targets the methodological and learning framework as well as creation of many tools based on free softwares 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

1 Introduction

1.1 A short description of literature review

Detailed literature review for pharmaceutical and medical industry will be carried out. Following topics will allow us to get in-depth well-rounded coverage: we will start with the study of origins of Pharmaceutical Industry to get current perspective. Then we will analyze data regarding clinical Trials and Origin of Randomized Clinical Trials (RCT) which are considered a gold standard for evidence generation. Subsequently literature covering modern Hospitals, Everyday Clinical Practice and Healthcare Environment will be studied. Insights from Real world evidence (RWE) and observational studies, use of Electronic Health Records (EHR) across the globe and evolution of statistics, and role of an analyst or programmer will be listed. A brief literature review for traditional Chinese Medicine history and philosophy will be provided. Next sub-sections cover the science of ayurveda, ongoing efforts at the national level by AYUSH and other ministries and development of modern medicine and outreach from an Indian context.

1.2 Science of ayurveda

There are some models proposed by various authors to handle complex and tricky situations arising in defining and understanding the action of mechanism of Ayurvedic intervention. As described by Dr. Girish Tillu in his talk at TDU, huge observational data for Ayurvedic medicines covers large number of books and manuscripts, 57 authentic books (Drug and cosmetic act 1940, page 47) [8], more than 4500 diseases including subtypes and conditions (Ayusoft database) [9], more than 81,000 formulations (TKDL database) [10], more than 4,00,000 Practitioners in India [11], infinite documents, references, experiential data, living tradition and knowledge in public domain [7]. Dravyaguna (Pharmacology), Bhaisajya Kalpana (Pharmaceutics), Nidana (Diagnosis) and Chikitsa (Management principles). This data points to a validated knowledge base.

Dr. D. B. Vaidya has explained the concept of reverse pharmacology to understand the action mechanism of Ayurvedic intervention. He says that there is huge amount of observational data available, showing relatively low side effects that have been reported. This existing data should be used as basis to carry out large interventional Ayurvedic trials to assess safety, efficacy, and pharmacokinetic information. This approach will be economical and could be less time consuming compared to the sequential drug development or the hierarchical model used in western medicine. He further talks about integrating meticulously documented experiential and experimental observations [12].

Prof R. H. Singh has opined that lab-based research experiments within Ayurvedic area during the last 50 years have not been rewarding. On the other hand, literary experiments to make a few of the classical Ayurvedic texts accessible to masses have been extremely useful. This situation warrants newer strategies of scientific research without compromising on the fundamental principles of Ayurveda [4].

Prof. Bhushan Patwardhan writes that there are substantial similarities between the traditional systems like Ayurveda and modern medicines. Ayurveda emphasizes on health promotion, disease prevention, early diagnosis, and personalized treatment. The modern medicine system approach uses predictive, preventive, and personalized medicine (PPPM). In case of Ayurveda, the evidence can be drawn from two main sources: (1) Evidence based on historical and classical nature of clinical practice supported by credible and accepted documentation. (2) Evidence based on ongoing scientific research to support various theories, medicines and procedures used in Ayurvedic medicine [13] [14].

Dr. Ram Manohar has expressed that Ayurveda is based on 5000 years of clinical practice. Hence, practice-based clinical trials should complement natural ways to gain insights [15].

Dr. Baghel's interpretation is that one should think of Ayurveda being in the developmental phase like any other medical systems. Like many other scholars he thinks that Ayurveda is a pure science based on logical explanation, which is called *Darshana*. Ongoing research in Ayurveda should impact academics, pharmacy, and practice in a profound way to convert data into information, information into knowledge and knowledge into wisdom [16].

As per Prof. Darshan Shankar's analysis as of 2015, at the national level, Ayurveda receives a meagre 3% of the Central health Budget and at the State level, making it difficult to fund any meaningful research projects. Despite Ayurveda's strengths, it has some limitations in current scenario. To advance the science there is a need to embrace tools of information technology to organize its vast multifaceted data, in searchable formats. Meticulously documented clinical experiences interpreted through Ayurveda-biology will expand rejuvenation of healthcare in India [17].

All the thought leaders cited here point to the strengths of Ayurveda as well as the immediate needs. They have pointed out that the research must be of high quality, and it must be impactful. They have indicated the need for experimental as well as experiential research. They have already provided a few new solutions and have urged to the research community to find new ways of tackling problems [18].

1.3 Potential opportunities for Real World Data analysis within Ayurveda

The western medicines are developed using a method called as hierarchical method where it tries answering questions with limited scope e.g., what is the efficacy of a particular drug, what is the safety profile of a drug? This method assumes a step wise approach and deals with the problem in successively conducted clinical trials of various types in a specific sequence. The pharmacology of the molecule is ascertained first at the very beginning. These studies are followed by cohort studies including open-label randomized studies but in general the clinical trial testing usually concludes with a blinded, randomized controlled trial (RCT). As already pointed out, RCTs are the "gold standard" of evidence generation as they offer most internal validity and minimal bias [19]. These studies could be complemented by using case studies, case series. This "one step at a time" approach has worked very well in the western medicine framework.

Ayurveda has been practiced for more than a millennium and is widely accepted in India as a worthy medical system. Over the last few decades' people across the world have gained knowledge and realized the importance of the age-old medical system and are constantly driven towards it. Although, having been in practice for ages Ayurveda still does not enjoy the recognition which the Western medicine does. Hence there needs to be a structured approach towards making this possible. The untapped potential of Ayurveda needs to be scientifically communicated globally for a wider reach for it to be utilized as a public health tool for promotion of health and prevention of diseases [19].

Ayurvedic vaidya 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 [11], 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.

1.4 What this study aims to contribute to

This study aims to contribute to interests of multiple stakeholders involved in clinical research. Based on arguments given above about availability of observational data, availability of technology, the Indian context and the specific case about Ayurveda and need to understand the science behind it, the study would highlight multiple use cases. The study would highlight tools and information generated through these tools and this could be interchangeably used by interested stakeholders including:

1.4.1 Hospital management

To keep any hospital functioning smoothly, operational insights from routine hospital data are important to improve management and efficiency of day-to-day activities. How many patients are present in the database? What are the characteristics of these patients? What is the gender distribution and age group distribution? Which countries, states, cities do they come from? How many times do they visit the hospital? What is the number of In-Patients & Out-Patients? What kind of assessments are done at each visit? What is the duration of visits for a patient? Which diseases are getting treated? Are the patients benefiting? How do you measure benefit?

Regular analysis of data would give insights which will allow for operational proficiencies, cost savings and eventually profitability. Finally, patients satisfaction would improve if a hospital functions efficiently.

1.4.2 Clinicians or treating doctors

Let us understand what kind of benefits the clinicians can have from this study. Traditionally many clinicians in their own practice use paper CRF to capture patient data. The hospital has electronic data capture system, which is essentially doing the same job, but data capture is electronically done. Are they ready to adopt to a new way of working? Does onboarding training at hospital cover this part of the job? Do they see this as an additional burden or an integral part of day-to-day work? How can experience of clinicians using the Health Information system be viewed? Do doctors like data entry part of job? Do the doctors have time for real data entry while consulting patients?

Answers to some of these questions can be generated indirectly by understanding the quality of the database and data contents. Does hospital management take any feedback from the clinicians who are the primary "end users" of the data capture system? This timely feedback loop should provide great inputs into evolution of the system both operationally as well as scientifically.

The electronic data capture provides unique opportunities to clinicians such as they have access to the data from other practicing clinicians. This gives indirect learning opportunities of understanding treatment protocols, treatment variations employed, rare diseases treated at the hospital. Retrospective analysis of disease and treatment should provide ideas about disease variations, appropriateness of documentation, disease – disease combinations, disease – treatment combinations. These documented combinations could be clinically meaningful, could be season wise, gender wise, age wise varying. Retrospective analysis of treatments should provide tendencies of treatment prescription such as use of classical treatments, herbo mineral treatments.

Data review and analysis tools developed for this thesis can enhance patient and doctor interactions.

1.4.3 Universities and students

Teaching material for students: Over the years, teaching methodology has not transformed even though there are quite a lot of advances in modern methodology. Can learning objectives of different kinds be tackled by making insights generated from this data available to the students. For example, can complex disease-treatment relationships be made easily visually available. Can interesting ways of explaining text and advice from Ayurveda which have contemporary relevance be created? For example, occurrence of certain diseases in certain geographic areas or season. Very few large-scale studies like this have taken place in Ayurveda, so this study with large amounts of observational data can provide exploratory opportunities to describe findings, textually as well as diagrammatically. Tools developed here can be used as a supplementary material for any MD / PhD student.

Scientific literature generation by researchers: Most hospitals in India or any part of the world mainly focus on treatment and not on research publications. Can a research team be put together for medical communication who publish papers as their primary job? If there is no known profile of patients visiting an Ayurvedic hospital and if this data can be generated and represented in the right form, it will provide novel information. How can we measure the strengths and weaknesses of an Ayurvedic practice? How should researchers evaluate changes in results of a practice over time? Is it possible to build new hypothesis? Is prescribed treatment truly personalized? Is it possible to trace back the treatment regimen followed and compare it to classical fundamentals? Is there a way to compare the demographics and patient characteristics from a Ayurvedic hospital against a mainstream western hospital? Which are rare diseases identified in the database? Can a clinically meaningful document be written, like a case series about this rare disease?

How can anyone use these data as "secondary use"? Can this data be used by insurance companies? Do the approved labels of medicines and prescriptions in the database match each other? Metal based formulations are questioned by non Ayurvedic community, what insights can be drawn about the rasa-aushadhis? Which are these medicines? For what diseases are they given and for what duration? Before providing the metal-based treatment and after providing the metal-based treatment, is there any difference in duration seen in treatments? What is the percentage of patients that are prescribed these medicines and what is the percentage of duration of all the duration of treatment given to these patients?

1.4.4 Policy makers – AYUSH and relevant ministries, insurance sector:

Policy makers and insurance companies can use insights from this data to decide on which treatments to cover for insurance. Government agencies can use this data to promote Ayurveda as a medical system throughout the world. They can make policies similar to policies that govern western medicine.

1.5 Introduction to real life data

As presented above, the potential benefits of gathering and analyzing such data are huge. A quick history of how this ability to collect data has come about is given below. This history also highlights the challenges of collecting and analyzing data in general on top of challenges associated with clinical research and Ayurveda specifically. With the passage of time, revolutions in technology have continually increased the creation of information and its exchange. With the advancement for communication from spoken to written, it became simpler to create texts, books thereby documentation; thus aiding transfer of knowledge from one person to another as well as from generation to generation without losing any data in translation. With increased and improved writing, compilation of articles, tables and records, there came a time where storing them became important, thus came in the libraries. The ability to effortlessly widen accumulated data had to wait until the 15th century. Around 1439, Johannes Gutenberg developed the printing press, causing an astonishing growth in the sharing of information at an economical cost.

The 20th century generated a remarkable growth in the publication of scientific journals and monographs, most of which were not critically reviewed, as most physicians had no way to access to the existing medical information. Towards the late 20th century, the spread of computers and the internet providing immediate virtual access to diverse information has entirely changed the way knowledge is collected, stored, and circulated. The flow of information has been increasing at almost exponential levels. Today, data sets are measured in zettabytes (10^21 bytes). Cost-effectively collected and stored data allows researchers across the world to successfully advance understanding of science and medicine [20].

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.

The relentless increase in the quantity and flood of information denotes an important professional opportunity, but a challenge simultaneously for those in medicine and science [21].

As indicated by Toby Cosgrove MD, from Cleveland, medical information doubles every 73 days, in the year 2020, as compared to approximately 3.5 years in 2010. An estimated 8,00,000 papers were published in 5,600 medical journals every year. It is projected that 12,000 new articles and 300 randomized controlled trials will be added to Medline each week, and that new medical articles will appear at a rate of one every 26 seconds [22]. To be able to generate any kind of analysis and make accurate predictions, there is a need to access, connect various sources, collate, and consume all the data. Data is being produced, obtained, and stored in numerous number of structures [23].

During an appointment at a hospital, diverse types of data are collected. Raw data are observations about individual patients created by the treating doctor at a hospital. These data may be in the form of measurements of patient's characteristics such as age, gender, height, weight, blood pressure, heart rate, etc. Raw data may also include description of the medical history, physical exam information, clinical laboratory results (e.g., serum lipid values, hemoglobin levels), whole exome or genome sequences, imaging results, ECGs, questionnaire data, or self-reported data (e.g., symptoms, quality of life).

Raw data, unprocessed source data, like unrefined gold buried deep in a mine is a precious resource. It is often: (1) Inconsistent, containing both relevant and irrelevant data, (2) Imprecise, containing incorrectly entered information or missing values, (3) Repetitive, containing duplicate data. To utilize the raw data to its fullest potential, it needs to be extracted, filtered through, understood, and transformed into analyzable format. One of the surveys carried out by Forbes estimates that data cleaning accounts for up to 80% of the development time and cost in data warehousing projects. Understanding the scope of data being analyzed and seeing the changes made to the data can accelerate the entire process of going from "information to building wisdom" [24] [25] [26] [27].

1.6 Introduction to clinical data understanding

The health care industry uses either a paper-based record keeping method and/or electronic health record (EHR) system to manage patient data. More and more organizations are using electronic data capture, but the practicing doctors in individual clinics may still be documenting observations on paper. The EHR has become an integral part of medical care, which transforms health care service quality and improves physicians' satisfaction and facilitates patients' decision. Accurate information from EHR enables physicians' decision making and measures clinical validity, which in turn upgrades the quality of patient care. This functionality is crucial during diagnosis and therapy, which benefits medical and legal practices too [23].

Health authorities and top-level journals require the data to be submitted along with research papers. The analyzable data set, is the result of many decisions made by varied people, as explained above. The errors, flaws, or biases in the processing of source data, will not necessarily be identified in the analyzable dataset. After the electronic data entry, new variables are generated to support further analysis. The final cleaned analyzable datasets consist of various components such as participant characteristics and primary outcome, pre-specified secondary and tertiary outcomes, adverse event data and exploratory data [28].

Physicians and other scientists are getting better at producing data. But we must become proficient—with or without the help of technology—at mining and managing the data in ways that will allow us to use it to maximum effect [29]. The full analyzable dataset is generally the most useful set of data to share, with large and likely important benefits to science and society. Secondly, the full analyzable dataset provides scientific validity to the outcome and ensures replication and repeatability. Further, meta-analysis increases the statistical power of detecting effects and maximizes the value of the outcome in the clinical knowledge base. Finally, analyzable data allows for further scientific discovery through additional secondary analyses, as well as the conduct of exploratory research to generate hypotheses for additional studies [30].

1.7 Introduction to study of demographics and patient characteristics

We will proceed with understanding the data contents and see if any of the questions raised earlier can be solved. Demography is the study of the population. It explains the composition, the distribution and the data trends seen in the population. Roles and functions for demography studies can be broadly defined as, (1) population projections, (2) inputs into government budget, (3) evidence-based policy, and (4) communication of vital statistics [31]. There is very little data on the profile of patients accessing traditional systems of medicine. A comparative study of profile of patients using an Ayurveda clinic and modern medicinal clinic will help in understanding of utilization of services and preference for health seeking behaviour.

1.8 Introduction to study of diagnostics and interventions

Diagnosis is one of the most important aspects in the process of treatment of the disease or condition. It is a patient-centered, cyclic process of gathering information, analyzing information, determining the health condition, and defining the type of intervention and continuously monitoring the progress till the desired state of functions/doshas is arrived at. Ayurvedic treatment involves removal of the causative factors. It assists in getting the functions/doshas into balance. The success of a treatment is possible only by timely and accurate diagnosis, a tailor-made intervention accompanied by an effective collaboration of the physician and the patient [32]. The term comorbidity refers to the coexistence of multiple diseases in relation to a primary disease in a patient. Patients report multiple diseases during their visits to the hospital. Some of these reported disorders are expected and some are unexpected. There are known as well as unknown disease combinations present due to biological linkages. Clinical and epidemiological studies indicate that disease comorbidities have a great impact on health status, selection of appropriate treatments and health system costs. Understanding comorbidities and their etiology is key to identify new preventive and therapeutic strategies.

Finally, all these steps (1) accessing and understanding real life data, (2) converting that data into analyzable format, (3) understanding demographics and patient characteristics, and (4) understanding diagnostics and interventions should help in building transdisciplinary evidence to increase the scientific understanding outside of the community, then increase the confidence and thereby widening the user base.

1.9 Structure of the thesis document

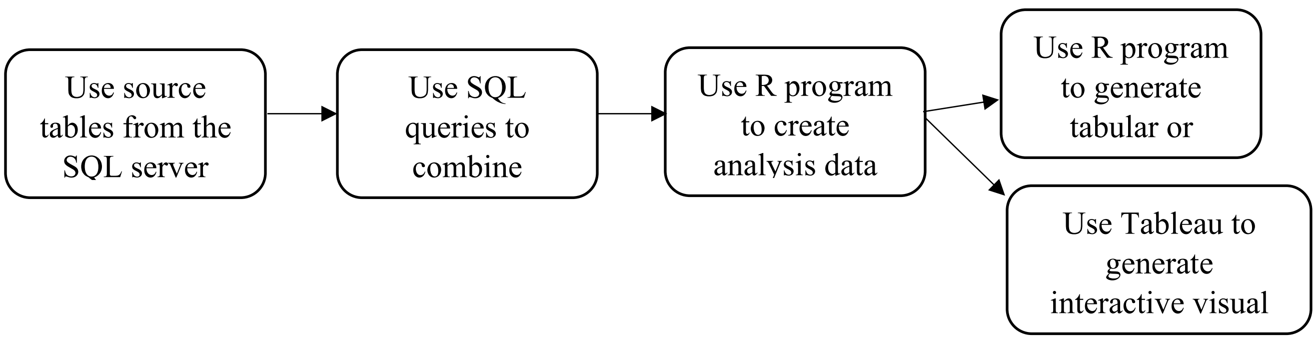
The thesis document will be structured as follows: Chapter 1 covers detailed literature review for pharmaceutical, medical industry and the need to carry out this study. Chapter 2 covers study design and numerous methods applied for data analysis. In section 2.7, table (Table *2-1*) presents a consolidated view of various analysis, context about different analysis and their possible relationships with each other. Detailed observations of these experiments are documented in chapter 3 Results. Challenges with respect to overall system architecture and data collection are documented in chapter 3. Chapter 4 provides some solutions to these challenges. Supplementary interpretations are further illustrated in chapter 4. Chapter 5 covers several conclusions drawn based on the analysis and key take away message. Chapter 6 appendix provides detailed material generated for this PhD work. Appendix section 6.2 provides details about the master dataset developed for whole of this study. Appendix section 6.3 provides information regarding the source database. Appendix section 6.4 provides a table to track all analysis presented in the thesis. Appendix section 6.5 provides R, SQL, and D3js programs developed to generate different datasets and analysis. Chapter 7 covers the references.

2 Methods and results

2.1 Study design

This was a retrospective study of Electronic Health Records (EHR) at TDU. Electronic Health Records of patients from 2011 to 2017 are used. It contained data for more than 51,000 patients, more than 1,50,000 visits, more than 900 variations of disease types, more than 3,000 variations of medical procedures. The study was approved by the authorities of IAIM and TDU (refer Appendix section 4.1). We explored "naturally reported data" for getting insights into demographics, health-seeking behaviors, and other health parameters. Sensitive information related to patients and doctors was not extracted to maintain confidentiality. Data was analyzed through SQL [33] and R programs [34], python [35], Java [36], D3js [37] and tableau [38] software. A high-level pictorial representation of the technical study is displayed below (Figure 2-1).

Figure 2-1: Pictorial representation of analysis



2.2 Data analysis design

The data analysis was represented using many different methods such as: (1) tabular representation using frequency counts [39], (2) descriptive summary statistics [39], (3) data representation on world / country map [40], (4) boxplot representation [41], (5) barplot [42] and (6) dotplot representation [42], (7) radar plot representation [43], (8) individual patient level data listings – line by line data representation, (9) various types of bubble plots [42], [44], (10) circular data representation [45], (11) collapsible tree diagram [46], (12) treemap / mosaic plot [47], (13) butterfly plot [48], (14) area plot [42], (15) calendar plot [42].

These various analyses enable data to be reported in different levels of details. Most of these representations are interactive, end user 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 [38].

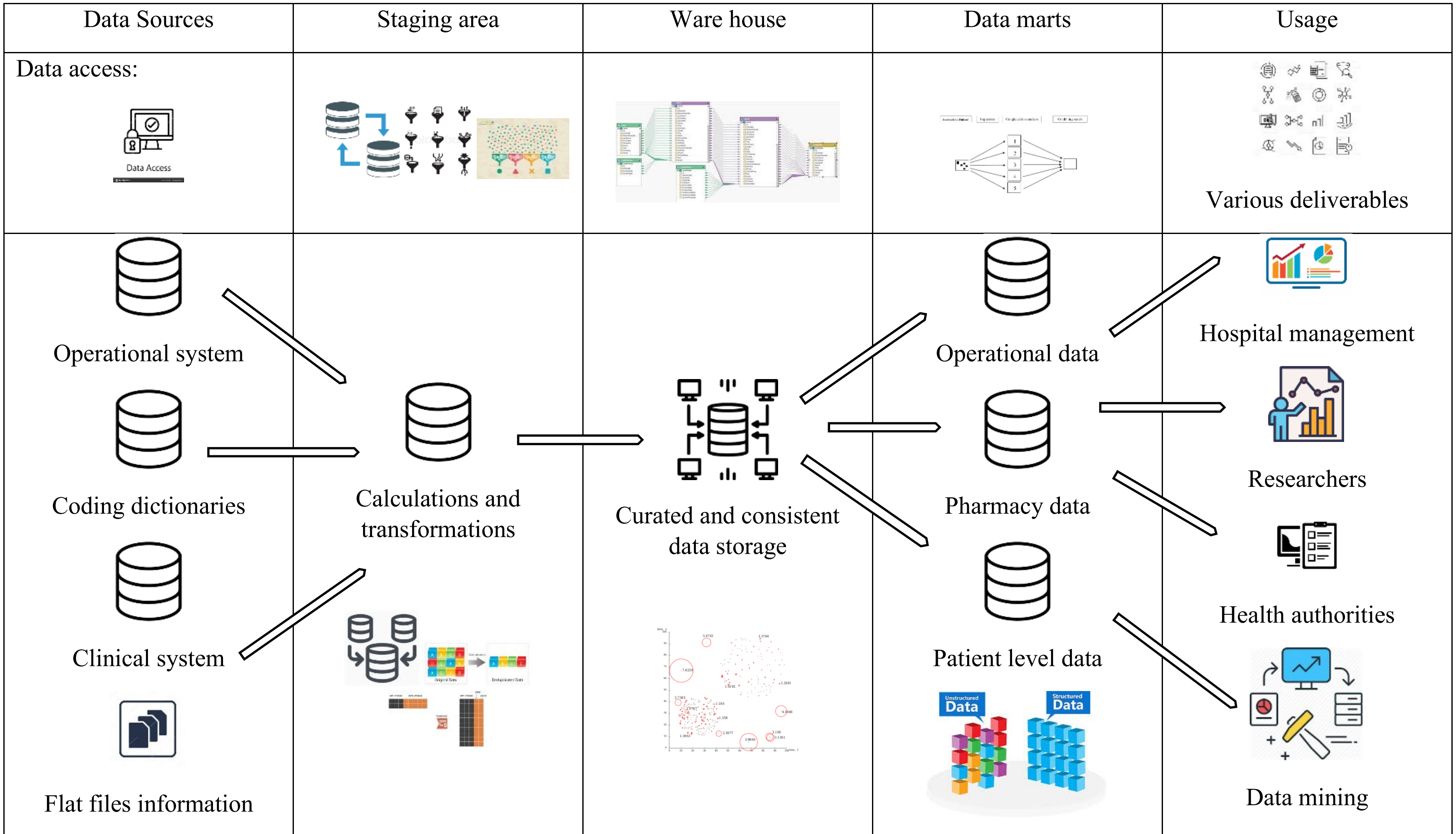
There are four major aspects involved in this study: (1) Converting real life clinical data into analyzable format: this includes accessing data, data preparation for analysis and derivation of additional information based on source data. Flow diagram from data source to final usage by various usage types is provided in (Figure *2-2*), (2) Clinical data understanding: this section covers broad checks on the datasets, contents checks, visit pattern analysis and patient disease treatment journey view, (3) Studying demographics and patient specific factors, and (4) Studying diagnosis and interventions.

The TDU / I-AIM team should be congratulated first before any discussion to create an electronic database right from the inception of the hospital. This foresight has allowed us to have significant amount of data. There are a lot of learnings from this exercise which can be beneficial to many institutes and hospitals.

Conversion of real-life clinical data from an individual data point to a logical dataset was done. Logical relationships were established post inspection of the datasets and the columns. Relational datasets were identified. Observations about data capture methods and storage were noted. Some shortcomings and errors in the data were seen and noted (missing data, inconsistent values, or unresolved duplicates). This exercise of understanding technical architecture from “an end user point of view” will help in running analysis of various types. If data generation for future use is one of the top priorities for the hospital, then there should be a project plan put together and appropriate steps should be taken to plug the existing gaps.

Results from these data explorations can be used by different stake holders like Hospital management (HM) from administration point of view, treating doctors (Medics), and basic researchers (Scientists). Table (Table *2-1*) systematically presents a consolidated view of various analysis, context about different analysis and their possible relationships with each other. In-depth explanation of methods, discussions, and results will be covered in the thesis document.

Figure 2-2: Flow diagram from data source to final usage by various usage types

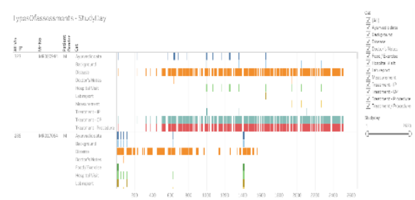
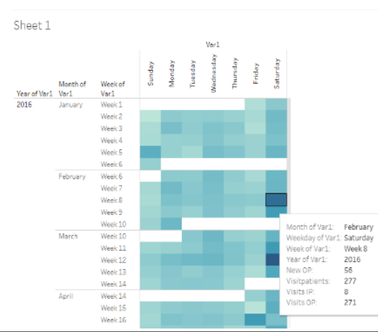


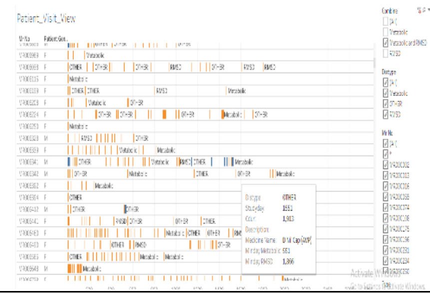
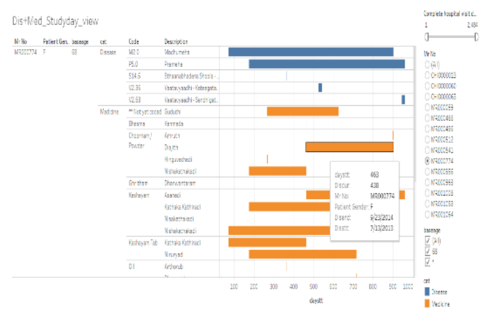
The source data from operational system, clinical system, and coding dictionaries is combined logically. Along with the source information, there are a few additional files created using subject matter expertise. These are logically added to the other source datasets. These give rise to intermediate 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.

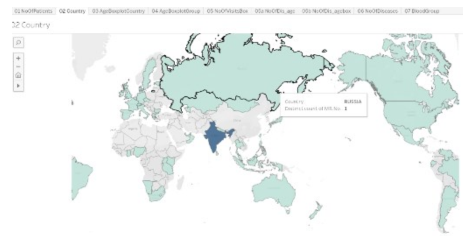
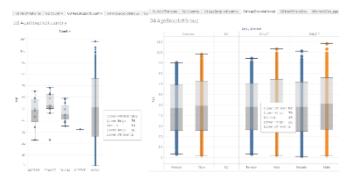
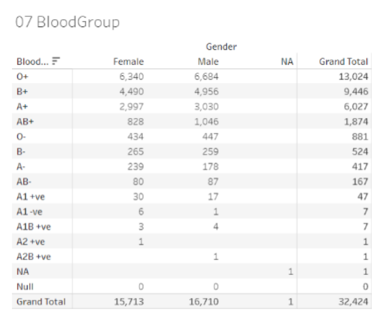
Final output datasets are stored in a central place called as a data warehouse. It is a central repository of data within for an organization. These datasets are used to make operational, business, and scientific decisions by various stakeholders by converting them into interactive analysis, dashboards, formal project reports.

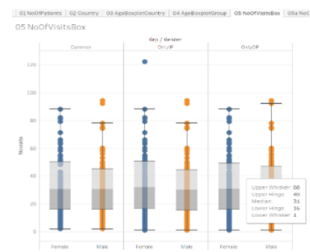
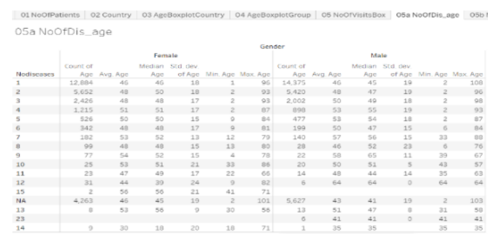
2.3 Summary of methods and results

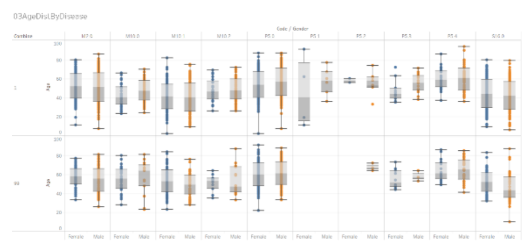
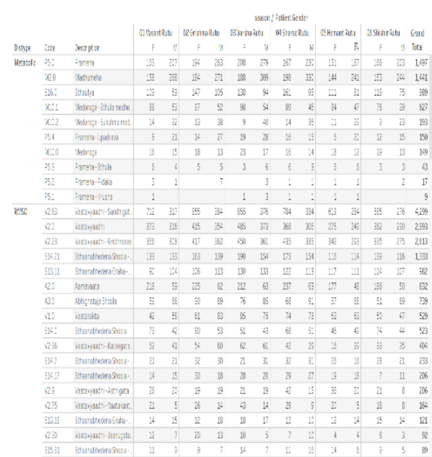
Table 2-1: Proposed methods and analysis use cases

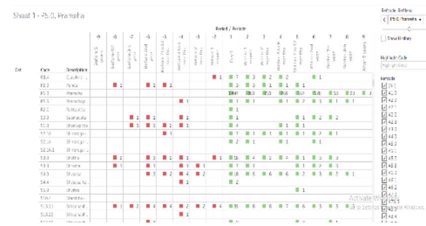
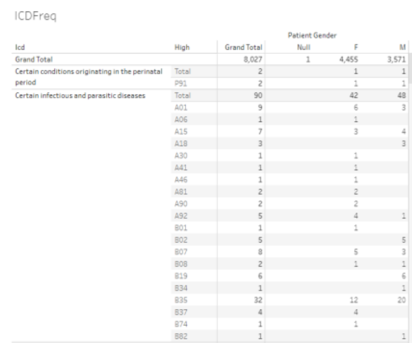
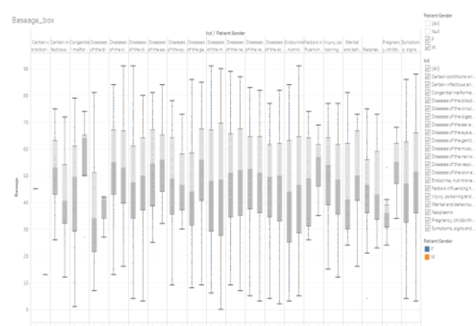
Stake holders	Classification	Type of analysis	Figure number and analysis name (linked to the actual figure)	Snapshot of the analysis (the picture is attached here to provide a quick look into the type of analysis)	Context of the proposed analysis
HM, Medics, Scientists	Clinical data understanding	Tabular frequency table	Figure 3-1: A snippet of disease table by gender		(1) Data quality check based on examples of a couple of diseases. (2) Generation of ideas about how “end users” are using the system
HM, Medics, Scientists	Clinical data understanding	Individual patient level data listing	Figure 3-2: Variable classification by categories		(1) Understanding data generation process at each visit for each patient to gain present status of data and provide feedback into improvements of system architecture. (2) Improve operational efficiencies of the Hospital Management Information system
HM	Clinical data understanding, operational efficiencies	Calendar plot	Figure 3-3: Visit pattern analysis		(1) Learn more about patient inflow and study how to increase outreach to the society. (2) Improve operational efficiencies across various departments based on patient visit patterns

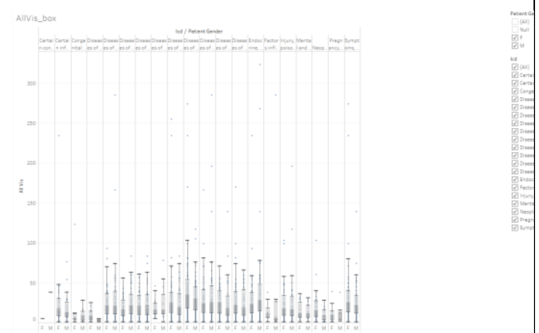
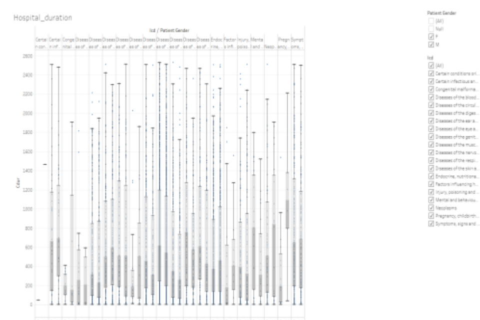
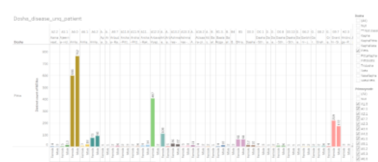
Stake holders	Classification	Type of analysis	Figure number and analysis name (linked to the actual figure)	Snapshot of the analysis (the picture is attached here to provide a quick look into the type of analysis)	Context of the proposed analysis
Medics	Patient disease and treatment journey	Individual patient level data listing	Figure 3-4: Patient visit profile – Horizontal view		(1) An analysis of how a patient journey is documented, how can this study be used prospectively and retrospectively to understand disease progression and treatment protocols (2) Convert treatment ideas into documented material (3) Improvements to the data standards and system architecture
Medics		Individual patient level data listing	Figure 3-5: Patient visit profile – Vertical view		
HM, Medics, Scientists	Study of demographics and patient	Tabular frequency table	Figure 3-6: Total Number of Patients	01 NoOfPatients | 02 Country | 03 AgeBoxplotCountry | 04 Age Summary of distinct number of patients Grp Common: 3,491; NA: 0; OnlyIP: 4,145; OnlyOP: 38,903; Grand Total: 39,557	(1) Demographic profiling analysis to understand health seeking behaviours of patients (2) Disease profiling based on preliminary disease analysis, visit patterns and types

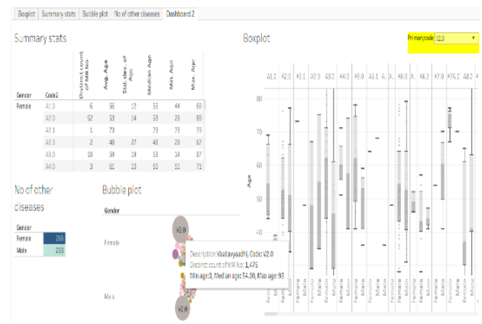
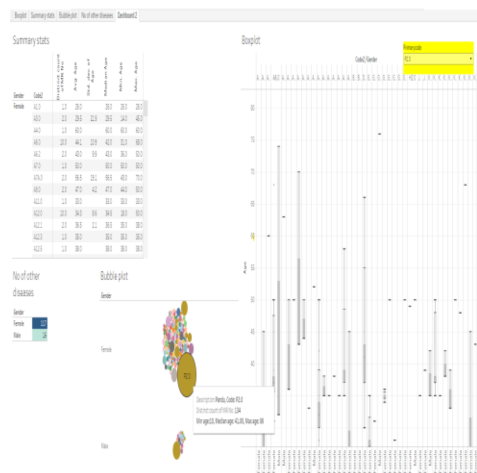
<table>
<tr><th>Stake holders</th><th>Classification</th><th>Type of analysis</th><th>Figure number and analysis name (linked to the actual figure)</th><th>Snapshot of the analysis (the picture is attached here to provide a quick look into the type of analysis)</th><th>Context of the proposed analysis</th></tr>
<tr><td>HM, Medics, Scientis ts</td><td rowspan="3">specific factors</td><td>Data on world map</td><td>Figure 3-7: Country-wise Visualization</td><td></td><td rowspan="3">of visits (In-patient and out-patient visits)
(3) Use such analysis to compare against demographic profiling of any other main-stream hospital to understand health seeking behaviours
(4) Use disease profiling study to understand epidemiology
(5) Use this analysis for defining public health policies to local and central authorities
(6) Secondary use of this analysis is to find data quality</td></tr>
<tr><td>HM, Medics, Scientis ts</td><td>Boxplot</td><td>Figure 3-8: Age distribution by country, age distribution by gender</td><td></td></tr>
<tr><td>HM, Medics, Scientis ts</td><td>Tabular frequenc y table</td><td>Figure 3-9: Blood-group Distribution by gender</td><td>
07 BloodGroup
<table>
<tr><th></th><th colspan="3">Gender</th><th></th></tr>
<tr><th>Blood...</th><th>Female</th><th>Male</th><th>NA</th><th>Grand Total</th></tr>
<tr><td>O+</td><td>6,340</td><td>6,684</td><td></td><td>13,024</td></tr>
<tr><td>B+</td><td>4,490</td><td>4,956</td><td></td><td>9,446</td></tr>
<tr><td>A+</td><td>2,997</td><td>3,030</td><td></td><td>6,027</td></tr>
<tr><td>AB+</td><td>828</td><td>1,046</td><td></td><td>1,874</td></tr>
<tr><td>O-</td><td>434</td><td>447</td><td></td><td>881</td></tr>
<tr><td>B-</td><td>265</td><td>259</td><td></td><td>524</td></tr>
<tr><td>A-</td><td>239</td><td>178</td><td></td><td>417</td></tr>
<tr><td>AB-</td><td>80</td><td>87</td><td></td><td>167</td></tr>
<tr><td>A1 +ve</td><td>30</td><td>17</td><td></td><td>47</td></tr>
<tr><td>A1 -ve</td><td>6</td><td>1</td><td></td><td>7</td></tr>
<tr><td>A1B +ve</td><td>3</td><td>4</td><td></td><td>7</td></tr>
<tr><td>A2 +ve</td><td>1</td><td></td><td></td><td>1</td></tr>
<tr><td>A2B +ve</td><td></td><td>1</td><td></td><td>1</td></tr>
<tr><td>NA</td><td></td><td></td><td>1</td><td>1</td></tr>
<tr><td>Null</td><td>0</td><td>0</td><td></td><td>0</td></tr>
<tr><td>Grand Total</td><td>15,713</td><td>16,710</td><td>1</td><td>32,424</td></tr>
</table></td></tr>
</table>

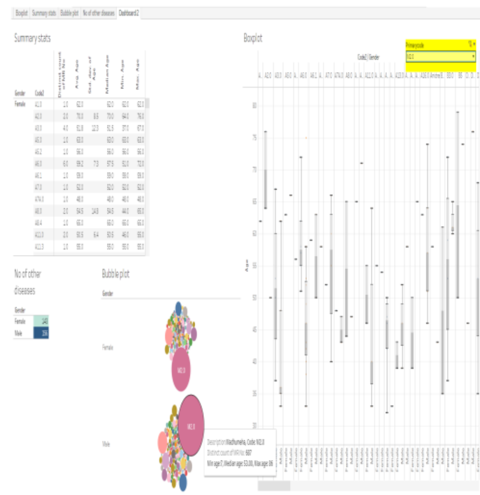
<table>
<tr><th>Stake holders</th><th>Classification</th><th>Type of analysis</th><th>Figure number and analysis name (linked to the actual figure)</th><th>Snapshot of the analysis (the picture is attached here to provide a quick look into the type of analysis)</th><th>Context of the proposed analysis</th></tr>
<tr><td>HM</td><td rowspan="2"></td><td>Boxplot</td><td>Figure 3-10: Number of Visits, and Visit Types</td><td></td><td rowspan="2"></td></tr>
<tr><td>HM, Medics, Scientists</td><td>Descriptive summary statistics</td><td>Figure 3-11: Descriptive summary statistics by number of Diseases by Age and Gender</td><td>
01 NoOfPatients | 02 Country | 03 AgeBoxplotCountry | 04 AgeBoxplotGroup | 05 NoOfVisitsBox | 05a NoOfDis_age | 05b N
05a NoOfDis_age
<table>
<tr><th></th><th colspan="12">Gender</th></tr>
<tr><th></th><th colspan="6">Female</th><th colspan="6">Male</th></tr>
<tr><th>Nodiseases</th><th>Count of Age</th><th>Avg. Age</th><th>Median Age</th><th>Std. dev. of Age</th><th>Min. Age</th><th>Max. Age</th><th>Count of Age</th><th>Avg. Age</th><th>Median Age</th><th>Std. dev. of Age</th><th>Min. Age</th><th>Max. Age</th></tr>
<tr><td>1</td><td>12,884</td><td>46</td><td>46</td><td>18</td><td>1</td><td>96</td><td>14,375</td><td>46</td><td>45</td><td>19</td><td>2</td><td>108</td></tr>
<tr><td>2</td><td>5,652</td><td>48</td><td>50</td><td>18</td><td>2</td><td>93</td><td>5,420</td><td>48</td><td>47</td><td>19</td><td>2</td><td>96</td></tr>
<tr><td>3</td><td>2,426</td><td>48</td><td>48</td><td>17</td><td>2</td><td>93</td><td>2,002</td><td>50</td><td>49</td><td>18</td><td>2</td><td>98</td></tr>
<tr><td>4</td><td>1,215</td><td>51</td><td>51</td><td>17</td><td>2</td><td>87</td><td>898</td><td>53</td><td>55</td><td>19</td><td>2</td><td>93</td></tr>
<tr><td>5</td><td>526</td><td>50</td><td>50</td><td>15</td><td>9</td><td>84</td><td>477</td><td>53</td><td>54</td><td>18</td><td>2</td><td>87</td></tr>
<tr><td>6</td><td>342</td><td>48</td><td>48</td><td>17</td><td>9</td><td>81</td><td>199</td><td>50</td><td>47</td><td>15</td><td>6</td><td>84</td></tr>
<tr><td>7</td><td>182</td><td>53</td><td>52</td><td>13</td><td>12</td><td>79</td><td>140</td><td>57</td><td>56</td><td>15</td><td>33</td><td>88</td></tr>
<tr><td>8</td><td>99</td><td>48</td><td>48</td><td>15</td><td>13</td><td>80</td><td>28</td><td>46</td><td>52</td><td>23</td><td>6</td><td>76</td></tr>
<tr><td>9</td><td>77</td><td>54</td><td>52</td><td>15</td><td>4</td><td>78</td><td>22</td><td>58</td><td>65</td><td>11</td><td>39</td><td>67</td></tr>
<tr><td>10</td><td>25</td><td>53</td><td>51</td><td>21</td><td>33</td><td>86</td><td>20</td><td>50</td><td>51</td><td>5</td><td>43</td><td>57</td></tr>
<tr><td>11</td><td>23</td><td>47</td><td>49</td><td>17</td><td>22</td><td>66</td><td>14</td><td>48</td><td>44</td><td>14</td><td>35</td><td>63</td></tr>
<tr><td>12</td><td>31</td><td>44</td><td>39</td><td>24</td><td>9</td><td>82</td><td>6</td><td>64</td><td>64</td><td>0</td><td>64</td><td>64</td></tr>
<tr><td>15</td><td>2</td><td>56</td><td>56</td><td>21</td><td>41</td><td>71</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>NA</td><td>4,263</td><td>46</td><td>45</td><td>19</td><td>2</td><td>101</td><td>5,627</td><td>43</td><td>41</td><td>19</td><td>2</td><td>103</td></tr>
<tr><td>13</td><td>8</td><td>53</td><td>56</td><td>9</td><td>30</td><td>56</td><td>13</td><td>51</td><td>47</td><td>8</td><td>31</td><td>58</td></tr>
<tr><td>23</td><td></td><td></td><td></td><td></td><td></td><td></td><td>6</td><td>41</td><td>41</td><td>0</td><td>41</td><td>41</td></tr>
<tr><td>14</td><td>9</td><td>30</td><td>18</td><td>20</td><td>18</td><td>71</td><td>1</td><td>35</td><td>35</td><td></td><td>35</td><td>35</td></tr>
</table></td></tr>
<tr><td>HM, Medics</td><td>Study of demographics and patient specific factors</td><td>Bubble plot</td><td>Figure 3-12: Data tabulation for patients reporting RMSD and Metabolic diseases</td><td></td><td>Analysis similar to described above on a subset of Metabolic and Rheumatic and Musculoskeletal disease (RMSD) patients</td></tr>
</table>

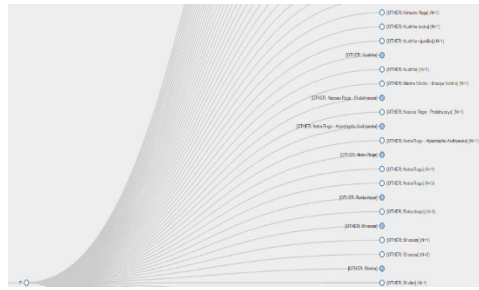
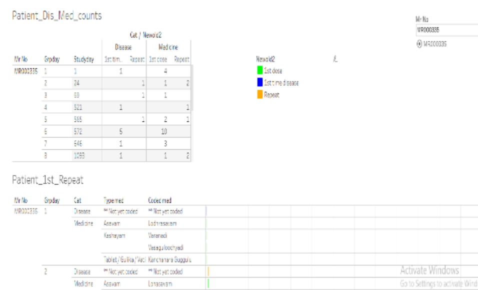
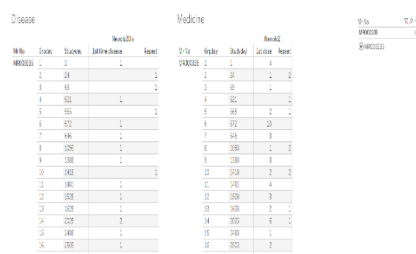
Stake holders	Classification	Type of analysis	Figure number and analysis name (linked to the actual figure)	Snapshot of the analysis (the picture is attached here to provide a quick look into the type of analysis)	Context of the proposed analysis
HM, Medics		Boxplot	Figure 3-13: Disease distribution by age and gender		
HM, Medics		Tabular frequency table	Figure 3-14: Patient visit duration for Disease categories by Gender		
HM, Medics	Disease pattern analysis	Tabular frequency table	Figure 3-15: Disease distribution by Seasonal Variations and gender		(1) Diseases are differently experienced by females and males as well as natural variations affect the prevalence – can we detect if the natural variations are reported in our database (2) A byproduct of this analysis is understanding of data entry and disease classification process at each visit

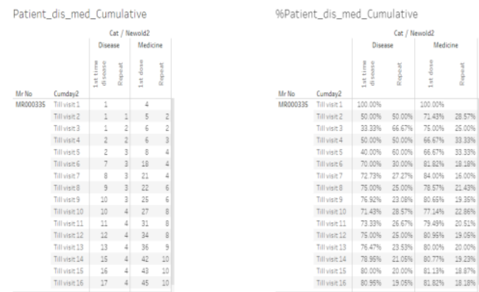
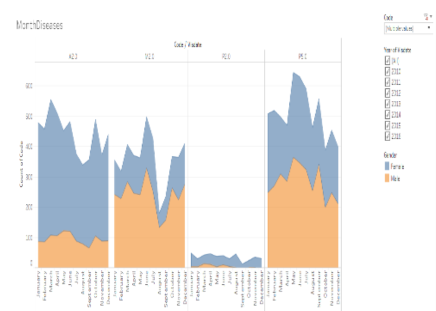
<table>
<tr><th>Stake holders</th><th>Classification</th><th>Type of analysis</th><th>Figure number and analysis name (linked to the actual figure)</th><th>Snapshot of the analysis (the picture is attached here to provide a quick look into the type of analysis)</th><th>Context of the proposed analysis</th></tr>
<tr><td>HM, Medics</td><td>Co-morbidity analysis</td><td>Tabular frequency table</td><td>Figure 3-16: Pre and Post Disease Classification Analysis</td><td></td><td>(1) Discover disease relationships by creating chronological view
(2) Differentiate pre and post diseases so that diagnosis and prognosis can be formed
(3) Pre and post differentiation is carried out for the prescribed medicines so that use of medicines could be understood at a summary level</td></tr>
<tr><td>HM, Medics, Scientists</td><td rowspan="2">Health seeking behaviour and public policy</td><td>Tabular frequency table</td><td>Figure 3-17: ICD classification by Gender</td><td>
ICDFreq
<table>
<tr><th>Icd</th><th>High</th><th>Grand Total</th><th>Patient Gender
Null</th><th>F</th><th>M</th></tr>
<tr><td>Grand Total</td><td></td><td>8,027</td><td>1</td><td>4,455</td><td>3,571</td></tr>
<tr><td>Certain conditions originating in the perinatal period</td><td>Total</td><td>2</td><td></td><td>1</td><td>1</td></tr>
<tr><td></td><td>P91</td><td>2</td><td></td><td>1</td><td>1</td></tr>
<tr><td>Certain infectious and parasitic diseases</td><td>Total</td><td>90</td><td></td><td>42</td><td>48</td></tr>
<tr><td></td><td>A01</td><td>9</td><td></td><td>6</td><td>3</td></tr>
<tr><td></td><td>A06</td><td>1</td><td></td><td>1</td><td></td></tr>
<tr><td></td><td>A15</td><td>7</td><td></td><td>3</td><td>4</td></tr>
<tr><td></td><td>A18</td><td>3</td><td></td><td></td><td>3</td></tr>
<tr><td></td><td>A30</td><td>1</td><td></td><td>1</td><td></td></tr>
<tr><td></td><td>A41</td><td>1</td><td></td><td>1</td><td></td></tr>
<tr><td></td><td>A46</td><td>1</td><td></td><td>1</td><td></td></tr>
<tr><td></td><td>A81</td><td>2</td><td></td><td>2</td><td></td></tr>
<tr><td></td><td>A90</td><td>2</td><td></td><td>2</td><td></td></tr>
<tr><td></td><td>A92</td><td>5</td><td></td><td>4</td><td>1</td></tr>
<tr><td></td><td>B01</td><td>1</td><td></td><td>1</td><td></td></tr>
<tr><td></td><td>B02</td><td>5</td><td></td><td></td><td>5</td></tr>
<tr><td></td><td>B07</td><td>8</td><td></td><td>5</td><td>3</td></tr>
<tr><td></td><td>B08</td><td>2</td><td></td><td>1</td><td>1</td></tr>
<tr><td></td><td>B19</td><td>6</td><td></td><td></td><td>6</td></tr>
<tr><td></td><td>B34</td><td>1</td><td></td><td></td><td>1</td></tr>
<tr><td></td><td>B35</td><td>32</td><td></td><td>12</td><td>20</td></tr>
<tr><td></td><td>B37</td><td>4</td><td></td><td>4</td><td></td></tr>
<tr><td></td><td>B74</td><td>1</td><td></td><td>1</td><td></td></tr>
<tr><td></td><td>B82</td><td>1</td><td></td><td></td><td>1</td></tr>
</table></td><td rowspan="2">(1) ICD classification of disease is used globally to understand the disease burden. What patterns emerge via this analysis, what health seeking behaviors can be understood?</td></tr>
<tr><td>HM, Medics, Scientists</td><td>Boxplot</td><td>Figure 3-18: Age distribution by ICD classification and Gender</td><td>
</td></tr>
</table>

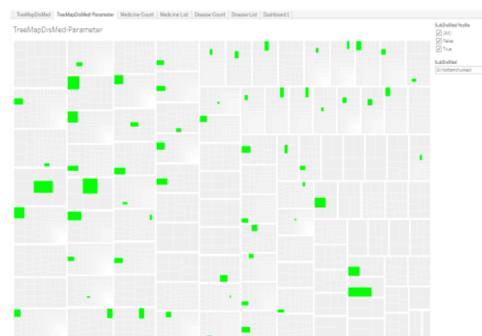
Stake holders	Classification	Type of analysis	Figure number and analysis name (linked to the actual figure)	Snapshot of the analysis (the picture is attached here to provide a quick look into the type of analysis)	Context of the proposed analysis
HM, Medics, Scientists		Boxplot	Figure 3-19: Visit distribution by ICD classification and Gender		
HM, Medics, Scientists		Boxplot	Figure 3-20: Duration distribution by ICD classification and Gender		
Medics, Scientists	Disease and Prakriti analysis	Barplot	Figure 3-21: Disease classification by Prakriti and Gender		What does Prakriti and disease data

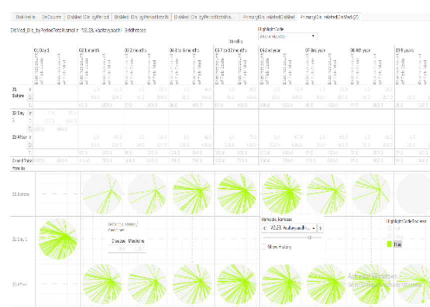
Stake holders	Classification	Type of analysis	Figure number and analysis name (linked to the actual figure)	Snapshot of the analysis (the picture is attached here to provide a quick look into the type of analysis)	Context of the proposed analysis
Scientists	Co-morbidity analysis	Bubble plot + Boxplot + Descriptive summary statistics	Figure 3-22: Co-morbidity analysis approach 1 example 1: Vaatavyadhi		Disease co-morbidities generate insights
Scientists		Bubble plot + Boxplot + Descriptive summary statistics	Figure 3-23: Co-morbidity analysis approach 1 example 2: Pandu		

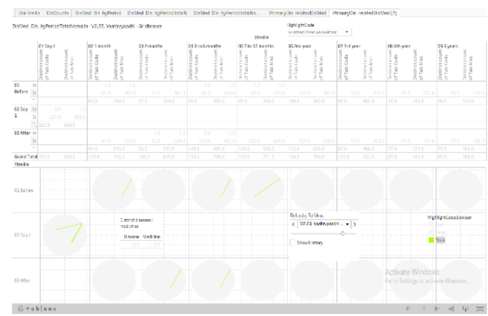
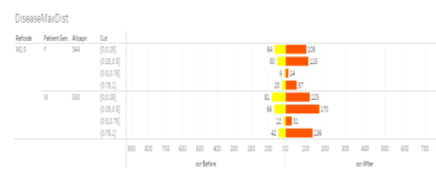
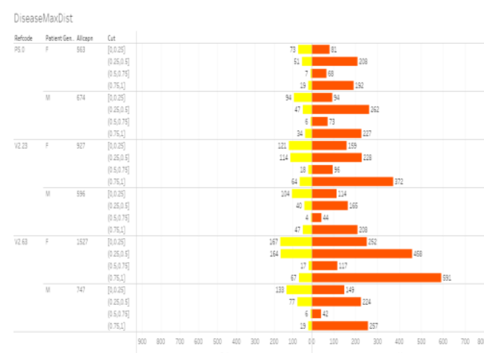
Stake holders	Classification	Type of analysis	Figure number and analysis name (linked to the actual figure)	Snapshot of the analysis (the picture is attached here to provide a quick look into the type of analysis)	Context of the proposed analysis
Scientists		Bubble plot + Boxplot + Descriptive summary statistics	Figure 3-24: Co-morbidity analysis approach 1 example 3: Madhumeha		
Scientists		Bubble plot + Tabular frequency table	Figure 3-25: Co-morbidity analysis approach 2		

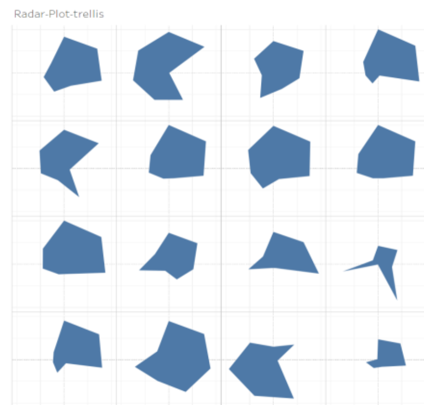
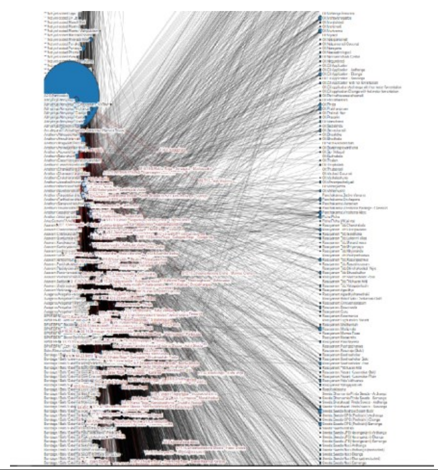
Stake holders	Classification	Type of analysis	Figure number and analysis name (linked to the actual figure)	Snapshot of the analysis (the picture is attached here to provide a quick look into the type of analysis)	Context of the proposed analysis
Medics, Scientists		Collapsible tree diagram	Figure 3-26: Co-morbidity analysis approach 3: collapsible tree view		
Medics	Patient disease and treatment journey	Individual patient level data listing + Tabular frequency table	Figure 3-27: Patient Disease and Treatment administration by Study Day		(1) An analysis of how a patient journey is documented, how can this study be used prospectively and retrospectively to understand disease progression and treatment protocols (2) This could help study the severity of the disease, co-morbidities and the number of medications prescribed to treat the condition (3) This can also provide an overview of the practicing physician's style of treatment and may be help draw parallels in treating medical conditions (4) Improvements to the data standards and system architecture is a byproduct of this analysis
Medics		Individual patient level data listing + Tabular frequency table	Figure 3-28: Patient Disease by Study Day and Treatment administration by Study Day		

Stake holders	Classification	Type of analysis	Figure number and analysis name (linked to the actual figure)	Snapshot of the analysis (the picture is attached here to provide a quick look into the type of analysis)	Context of the proposed analysis
Medics		Individual patient level data listing + Tabular frequency table	Figure 3-29: Patient Cumulative Disease and Treatment administration by Visit		
Medics, Scientists	Disease pattern analysis	Area plot	Figure 3-30: Area graph representation of diseases		(1) Due to this analysis representation, information on many diseases can be viewed in very short space (diseases plotted side by side) (2) Diseases vary by seasons, by gender as well as some diseases are more prevalent than others which can be seen, and clinical interpretations can be drawn (3) Operational and clinical insights can be generated with help of this visualization
Medics, Scientists	Co-morbidities and concomitant medications	Mosaic plot	Figure 3-31: Mosaic plot: Disease and treatment representation example 1: Prameha		(1) In ayurveda, a treatment is used for multiple diseases and multiple treatments are used for the same disease based on the context of disease. This relationship gives rise to many to many associations between disease and treatment (2) Many to many relationships which are hard to visualize are generated through this analysis

<table>
<tr><th>Stake holders</th><th>Classification</th><th>Type of analysis</th><th>Figure number and analysis name (linked to the actual figure)</th><th>Snapshot of the analysis (the picture is attached here to provide a quick look into the type of analysis)</th><th>Context of the proposed analysis</th></tr>
<tr><td>Medics, Scientists</td><td rowspan="4"></td><td>Tabular frequency table</td><td>Figure 3-32: Disease and treatment example 2: P5.0: Prameha and Oil: Kottamchukkadi</td><td>Medicine-List<table><tr><th>Med02</th><th>Code02</th><th>totpatmed</th><th>totpatdis</th><th>n</th><th>percdis</th><th>percmed</th></tr><tr><td>Oil:Kottamchukkadi</td><td>Metabolic:P5.0:Prameha</td><td>4604</td><td>1456</td><td>324</td><td>22.25</td><td>7.04</td></tr></table></td><td rowspan="4">(3) This analysis produces a mosaic display for diseases showing what kinds of treatments are prescribed
(4) Another mosaic display is created for treatments showing what kinds of diseases are getting treated by the underlying treatment
(5) Clinically meaningful as well as not so meaningful relationships are visualized, rare disease – disease combinations or disease – medicine combinations also can be studied</td></tr>
<tr><td>Medics, Scientists</td><td>Tabular frequency table</td><td>Figure 3-33: Disease and treatment example 3: P5.0: Prameha and Vati: Diabecon DS</td><td><table><tr><td>Tablet / Gulika / Vati:Diabecon DS</td><td>Metabolic:P5.0:Prameha</td><td>476</td><td>1456</td><td>145</td><td>9.96</td><td>30.46</td></tr></table></td></tr>
<tr><td>Medics, Scientists</td><td>Mosaic plot</td><td>Figure 3-34: Mosaic plot Disease and treatment representation example 4: Treatment: Oil: Kottamchukkadi</td><td></td></tr>
<tr><td>Medics, Scientists</td><td>Tabular frequency table</td><td>Figure 3-35: Cross tabulation of prescribed treatments and disease group by gender Example 1</td><td>MedType_DisType<table>
<tr><th></th><th colspan="7">Distype / Patient Gender</th></tr>
<tr><th></th><th colspan="3">Metabolic</th><th colspan="3">RMSD</th><th rowspan="2">Grand Total</th></tr>
<tr><th>Medicine Name</th><th>F</th><th>M</th><th>Total</th><th>F</th><th>M</th><th>Total</th></tr>
<tr><td>Balarishtam</td><td></td><td></td><td></td><td>29</td><td>10</td><td>39</td><td>39</td></tr>
<tr><td>Balarishtam 450 ml</td><td>3</td><td>2</td><td>5</td><td>143</td><td>80</td><td>223</td><td>225</td></tr>
<tr><td>Balarishtam 450 ml [sna]</td><td></td><td></td><td></td><td>16</td><td>9</td><td>25</td><td>25</td></tr>
<tr><td>Balarishtam 450 ml KPL</td><td></td><td></td><td></td><td>13</td><td>9</td><td>22</td><td>22</td></tr>
<tr><td>Balarishtam 450 ML SNA</td><td></td><td></td><td></td><td>9</td><td>5</td><td>14</td><td>14</td></tr>
<tr><td>Balarishtam 450 ml Vaidya</td><td>3</td><td>2</td><td>5</td><td>158</td><td>122</td><td>280</td><td>282</td></tr>
<tr><td>Balarishtam [KPL]</td><td></td><td></td><td></td><td>2</td><td>1</td><td>3</td><td>3</td></tr>
<tr><td>Balarishtam [AVP]</td><td>11</td><td>8</td><td>19</td><td>353</td><td>256</td><td>609</td><td>616</td></tr>
<tr><td>Balarishtam Rajah</td><td></td><td></td><td></td><td>1</td><td>1</td><td>2</td><td>2</td></tr>
<tr><td>Balarishtam(Arya)</td><td>1</td><td>1</td><td>2</td><td>24</td><td>12</td><td>36</td><td>37</td></tr>
<tr><td>Grand Total</td><td>17</td><td>13</td><td>30</td><td>677</td><td>465</td><td>1,142</td><td>1,154</td></tr>
</table></td></tr>
</table>

<table>
<tr><th>Stake holders</th><th>Classification</th><th>Type of analysis</th><th>Figure number and analysis name (linked to the actual figure)</th><th>Snapshot of the analysis (the picture is attached here to provide a quick look into the type of analysis)</th><th>Context of the proposed analysis</th></tr>
<tr><td>Medics, Scientists</td><td></td><td>Tabular frequency table</td><td>Figure 3-36: Cross tabulation of prescribed treatments and disease group by gender Example 2</td><td>MedType_DisType
<table>
<tr><th></th><th colspan="7">Distype / Patient Gender</th></tr>
<tr><th></th><th colspan="3">Metabolic</th><th colspan="3">RMSD</th><th>Grand</th></tr>
<tr><th>Medicine Name</th><th>F</th><th>M</th><th>Total</th><th>F</th><th>M</th><th>Total</th><th>Total</th></tr>
<tr><td>Abhrak bhasma (Dooth) 2gm</td><td></td><td></td><td></td><td>2</td><td>1</td><td>3</td><td>3</td></tr>
<tr><td>Abhraka Bhasma 1g (Sahasraputi) Dhoot</td><td></td><td></td><td></td><td></td><td>1</td><td>1</td><td>1</td></tr>
<tr><td>Abhraka Bhasma 5g [101] (AVP)</td><td>3</td><td>2</td><td>5</td><td>1</td><td>2</td><td>3</td><td>8</td></tr>
<tr><td>Abhraka Bhasma AVP</td><td>3</td><td>2</td><td>5</td><td>2</td><td>1</td><td>3</td><td>8</td></tr>
<tr><td>Abhraka Bhasma Dooth-100Butas</td><td>1</td><td>2</td><td>3</td><td>2</td><td>6</td><td>8</td><td>10</td></tr>
<tr><td>Abhrakam bhasma</td><td></td><td></td><td></td><td>2</td><td></td><td>2</td><td>2</td></tr>
<tr><td>Apamarga kshara Bhasma</td><td></td><td></td><td></td><td></td><td>1</td><td>1</td><td>1</td></tr>
<tr><td>Godhanti Bhasma 50gm [SDM]</td><td>1</td><td></td><td>1</td><td>16</td><td>4</td><td>20</td><td>20</td></tr>
<tr><td>Hajrulayahda Bhasma (Baidyanath)</td><td>1</td><td>5</td><td>6</td><td>3</td><td>8</td><td>11</td><td>17</td></tr>
<tr><td>Heeraka Bhasma 100mg [Dooth]</td><td></td><td></td><td></td><td>2</td><td></td><td>2</td><td>2</td></tr>
<tr><td>Kanmada Bhasma(10gm) (AVP)</td><td>4</td><td>4</td><td>8</td><td>2</td><td>2</td><td>4</td><td>11</td></tr>
<tr><td>Kashis Bhasma (Baidyanath)</td><td></td><td></td><td></td><td>1</td><td></td><td>1</td><td>1</td></tr>
<tr><td>Pravala bhasma</td><td>1</td><td></td><td>1</td><td>2</td><td>2</td><td>4</td><td>5</td></tr>
<tr><td>Pravala Bhasma Tab Dooth</td><td>1</td><td></td><td>1</td><td></td><td>1</td><td>1</td><td>2</td></tr>
<tr><td>Pravala Bhasmam (AVP)</td><td>1</td><td></td><td>1</td><td>1</td><td></td><td>1</td><td>2</td></tr>
<tr><td>Pravala Bhasmam Cap - 100</td><td></td><td></td><td></td><td>10</td><td>2</td><td>12</td><td>12</td></tr>
<tr><td>Pravala Bhasmam Cap (Kottakkal)</td><td>2</td><td></td><td>2</td><td>98</td><td>35</td><td>133</td><td>133</td></tr>
<tr><td>Rajatabhasma Cap (Kottakkal)</td><td>1</td><td></td><td>1</td><td></td><td></td><td></td><td>1</td></tr>
<tr><td>Roupya (Rajatha) Bhasma</td><td></td><td></td><td></td><td></td><td>1</td><td>1</td><td>1</td></tr>
<tr><td>Roupya (Rajatha) Bhasma 5 gm (Dooth)</td><td></td><td></td><td></td><td>2</td><td>1</td><td>3</td><td>3</td></tr>
<tr><td>Sameerapannaga Bhasma (Dooth)</td><td></td><td></td><td></td><td>2</td><td></td><td>2</td><td>2</td></tr>
<tr><td>sanka bhasmam</td><td></td><td></td><td></td><td>2</td><td></td><td>2</td><td>2</td></tr>
<tr><td>sankha bhasma</td><td></td><td></td><td></td><td>1</td><td>1</td><td>2</td><td>2</td></tr>
<tr><td>Sankha Bhasma(AVP)</td><td></td><td></td><td></td><td>1</td><td></td><td>1</td><td>1</td></tr>
<tr><td>Shilajith Bhasma</td><td>1</td><td>2</td><td>3</td><td></td><td></td><td></td><td>3</td></tr>
<tr><td>Shuddha Shilajit Bhasma</td><td></td><td>1</td><td>1</td><td></td><td></td><td></td><td>1</td></tr>
<tr><td>Tankan Bhasma 10gm</td><td></td><td>1</td><td>1</td><td>1</td><td>1</td><td>2</td><td>3</td></tr>
<tr><td>Tankan Bhasma (Baidyanath)</td><td></td><td>2</td><td>2</td><td>3</td><td>2</td><td>5</td><td>7</td></tr>
<tr><td>Tapyadi Loha Bhasma (Dooth)</td><td>1</td><td></td><td>1</td><td>3</td><td>1</td><td>4</td><td>4</td></tr>
<tr><td>Trivanga Bhasma (Dooth)</td><td>1</td><td>1</td><td>2</td><td></td><td></td><td></td><td>2</td></tr>
<tr><td>Vanga Bhasma 5g (Dhoot)</td><td>1</td><td>2</td><td>3</td><td></td><td></td><td></td><td>3</td></tr>
<tr><td>Yashada Bhasma 5gm (Dooth)</td><td>3</td><td>6</td><td>9</td><td>10</td><td>12</td><td>22</td><td>28</td></tr>
<tr><td>Grand Total</td><td>24</td><td>30</td><td>54</td><td>161</td><td>82</td><td>243</td><td>287</td></tr>
</table></td><td></td></tr>
<tr><td>Medics, Scientists</td><td>Co-morbidities and concomitant medications</td><td>Circular data representation</td><td>Figure 3-37: Circular view: Co-occurrences of disease – disease Example 1</td><td></td><td>(1) This analysis provides a view for a disease – disease combination or disease – medicine combination longitudinally (day 1 of disease, diseases reported at different time points before and after day 1)
(2) Disease and treatment information is represented on a circular display with green spokes representing co-occurrences of disease combination or disease – medicine</td></tr>
</table>

Stake holders	Classification	Type of analysis	Figure number and analysis name (linked to the actual figure)	Snapshot of the analysis (the picture is attached here to provide a quick look into the type of analysis)	Context of the proposed analysis
Medics, Scientists		Circular data representation	Figure 3-38: Circular view: Co-occurrences of disease – treatment Example 2		(3) Clinically meaningful as well as not so meaningful relationships are visualized, rare disease – disease combinations or disease – medicine combinations also can be studied
Scientists	Similarity analysis in disease and medicine trajectories	Butterfly plot	Figure 3-39: Pre and Post distance analysis for disease: M2.0: Madhumeha		(1) Disease trajectories (chronological list of diseases reported) and medicine trajectories (chronological list of medicines prescribed) are generated to study if similar diseases are experienced after an onset of a particular disease (2) This analysis helps in understanding biological changes, represented by subsequent reported diseases triggered by an underlying disease (3) Similarly, which medicines are prescribed provide insights into treatment protocols
Scientists		Butterfly plot	Figure 3-40: Pre and Post distance analysis for medicines given for diseases: P5.0, V2.23, V2.63		

Stake holders	Classification	Type of analysis	Figure number and analysis name (linked to the actual figure)	Snapshot of the analysis (the picture is attached here to provide a quick look into the type of analysis)	Context of the proposed analysis
Scientists	Multi-dimensional data analysis	Radar plot	Figure 3-41: Radar plot		(1) This analysis developed showed 7 parameters on 7 vertices. Different diseases were displayed next to each other as trellis radar display. (2) If there are different shapes for different diseases then this suggests that there are underlying differences in the data representing differences in diseases.
Scientists	Co-morbidities and concomitant medications	Dynamic bubble plot	Figure 3-42: Dynamic bubble plot: Example 1: Disease: A6.0: Amavaata		(1) Intricate relationship between disease and medicines in a very short space, at present this type of data representation approach may not have a direct application, but consultation with "end users" may provide appropriate use case.

3 Conclusion

The introduction of this thesis outlines the need to undertake such a study, by providing perspectives on medicine, pharmacy, development of hospitals throughout the world, internet era, and the Indian context relating to modern medicine as well as Ayurveda. Insights from the thought leaders in the field of Ayurveda are profound and they call for modern methods, new approaches and innovative strategies to be attempted to take Ayurveda science forward. There is a reflection on the type of evidence generated through controlled experiments (RCTs) and life experiments (Observational, Experiential) – some evidence about how these multiple approaches could yield meaningful results. This chapter asks a few questions from diverse points of views and seeks answers – some of these answers are hidden in everyday Ayurvedic clinical practice – which is still largely untapped, prompting the following: how can data analysis of electronically captured data help in advancing our understanding of Ayurveda as a practice and science?

Next section of this thesis elaborates on the technical details of a hospital HMIS based database and the associated EMR. We highlight technical details about the hospital database: how many source tables, how they stored in ~200+ tables, out of which ~20 to 25 tables were used to generate datasets useful for further analyses. Subsequently we display flowcharts outlining ~50+ 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. This will form the basis of possible operational and clinical analyses going forward.

Clinical data understanding section shows how individual observations can be transformed into meaningful patient narratives. This section explains how the usage of operational and clinical part of the data can benefit varied stake holders and emphasized the need to convert "a thought from a doctor's mind" into an "actionable and consistent data point" in the database for future use. While studying both the structure and the content of the hospital database, it was observed, that standardization of database along with effective curation towards good data quality is needed. It was highlighted that this type of data could provide a gold mine of information which when summarized could lead us to many insights which could potentially increase operational efficiencies and progress the Ayurvedic practice. It lays down the foundation for "understanding demographics and patient characteristics" as a basis.

The demographic and patient characteristic analysis provides good insights into different components of the data which can feed more generally into the public health domain. It is observed that the health and healthcare requirements of a population can benefit through disease surveillance and population health insights. It also provides actionable inputs to hospital management on efficient operations, practicing doctors and for research publications.

Diagnostics and interventions section shows complex relations between diseases and interventions. Comorbidities as well as combinations of interventions show the complex clinical decision-making process as followed by Ayurveda physicians. The disease and treatment comorbidity analyses are performed and presented using a variety of plots and heatmaps. These analyses show how individual observations can be transformed into meaningful insights and data stories.

Day-to-day transactions at hospitals involve people from many backgrounds like, hospital administration, patients, doctors, nurses, pharmacists, pathologists, representative from insurance companies, lawyers, etc. These interactions generate a lot of information and are the primary data generators. Same set of people and a few additional professionals are the end users of the data e.g., scientists, statisticians, database developers, etc. This study has provided preliminary insights into various aspects of HMIS based EMR data generated during real time consultation at I-AIM.

A variety of analysis and summarization of the hospital data are conducted with a view to derive meaningful outcomes which confirm the Ayurvedic principles. As a final summary of all that has been said previously is represented in the figure below (Figure *3-1*). First part of the diagram covers how to define the underlying question, which is followed by defining the hypothesis. Researchers should logically think about clinical context and methodological context. For converting this theoretical thinking into a real-world study, we need to have the necessary IT infrastructure which will enable data related components. Middle of the diagram shows potential stakeholders. This concise representation shows how the HMIS based EMR can shape up Real World Evidence generation.

Figure 3-1: Real World Evidence life cycle

Is the study specific to a treatment?	Is the treatment an approved one?	Is it a comparative study?	Has treatment been assigned by study protocol?	Is data available in existing sources?

Causal diagram: Specify causal relations & supporting evidence among treatment, outcome(s), & other variables to control confounding

				All Data Sources				
		Data protection	Different data types	IT infrastructure	Data quality	Information governance		
	Valid sample size	Clinical outcome	Disease registry	Patient registry	Patient charts	Sensor data, Mobile App	Low recall bias	
	Medical practitioner bias	Patient reported outcome		Real World Data		Longitudinal data	Co-morbidities and Cost effectiveness	
Methodology context	Low adherence	Quality of life outcome		Individual Patient Data (pragmatic trials, cohort trials, observational) Effectiveness in wider population		Health surveys	Preference of other medicine	Clinical context
	Confounding and Population homogeneity	Economic outcome		Stakeholders: Regulatory Authorities, Policy Makers, Government and Payers		Hospital EHRs	Real life data, clarity of treatment impact and AEs	
	Un-blinded treatment and Treatment switch	Primary / Secondary data	Retrospective / Prospective study	Big data, large sample size	Social media	Individual practice	More data available on drug and life style interaction	
		Operational challenges	Comparable data	Patient level data access	GDPR and Anonymization	Incomplete data		
				Data context				

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
    a. Sustainable treatment solutions for diseases readily available
    b. Thought provoking work to generate new needs through unconventional use of the data
    c. Expand the use of modern IT solutions like IT infrastructure, electronic health records, cloud, etc. within Ayurvedic area where appropriate – Ayur IT solutions.
    d. Take advantage of freely available cutting-edge software(s) to create new approaches
    e. Introduce statistical programming (Ayurdata analyst) as a tool to Ayurvedic area

A lot of work carried is out by health authorities, pharmaceutical companies, and nonprofit organizations. Some of these resources could be used as reference to become a world class data generator:

Clinical Data Interchange Standards Consortium (CDISC) is a global nonprofit charitable organization with administrative offices in Austin, Texas, with many people contributing across the world. CDISC brings experts together to create and advance data standards. This allows for accessibility, interoperability, and reusability of data for competent research that has greater impact on global health [49]. Many of the leading health authorities use the standards developed by the CDISC teams in various parts of drug applications.

TransCelerate BioPharma Inc. is a nonprofit organization with a mission to collaborate across the global biopharmaceutical research and development community to identify, prioritize, design, and facilitate the implementation of solutions designed to deliver high quality new medicines. They have many open-source solutions which could be used to improve delivery model [50].

Based on the 21st century act, 2016, the US FDA has created a regulatory framework for the Real-world data (RWD) and real-world evidence (RWE) which are playing increasing role in the health care decisions [51], [52]. The European Medicines Agency (EMA) has established a center to provide timely and reliable evidence from real world healthcare databases on the use, safety, and effectiveness of medicines for human use, including vaccines, across the European Union (EU). This capability is called the Data Analysis and Real-World Interrogation Network (DARWIN EU®) [53].

These resources provide a lot of material to enhance overall understanding and allows researchers to be compliant with the regulatory requirements.

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 *3-1*). We believe that this would be a pioneering effort within ayurvedic EMR area.

Table 3-1: Different types of documents

Medical Reporting	Medical Teaching	Medical advertising of products	Publication
• Newspaper & magazine articles • Mostly for public • Written in simple, non-technical language	For doctors • Textbooks, • Continued Medical Education programs, • Slide decks, • Online learning material • For Patients learning material	• Promotional information for healthcare professionals • Product profiles • Brochures • Sales force training • Online learning material	• Abstracts • Journal articles, case reports, review articles • Posters & presentations for scientific conferences

Research Documents	Regulatory documents #	
• Research proposals • Clinical trial protocols • Investigators' Brochure	• Package Inserts • Patient Information Leaflets • Clinical study reports	Aggregate safety reports such as • Periodic Safety Update Reports

• Informed Consent Documents • Study reports	• Web synopses • Subject narratives Common Technical Document (CTD) modules such as • nonclinical and clinical overviews & summaries • expert reports • PK, Safety, Efficacy summaries	• Periodic Adverse Drug Experience Reports • Annual safety reports • Policy papers

#: Presently, some of these documents are not applicable within Ayurvedic area

Work carried out for this thesis is typically reflective of work done by a team of people. In a mid to large sized pharmaceutical organization, this type of work is carried out by (1) Clinicians and statisticians design clinical protocol, (2) Database development team creates database and data flow components, (3) Data management team reviews and cleans the data on an ongoing basis, (4) Statistical programming team and statisticians create the necessary analyses, (5) Writing team generates Clinical Study Report / Publication, and last but not the least (6) IT team handles various systems so that the data and information flow is managed appropriately.

Much more details about the database and programming done for analyses and visualization are available in the appendices.

This is not an end but just a beginning of Ayurdata experts ...

4 Appendix

4.1 Approval from the hospital management to carry out the retrospective study

11.1 ANNEXURE 1 –NOC FROM IAIM MEDICAL DIRECTOR.

****

IAIM/2020/NOC/01 Date: 29.05.2020

LETTER OF PERMISSION AND NO OBJECTION CERTIFICATE

TO WHOM IT MAY CONCERN

This is to grant permission to Mr. Vinay Mahajan to conduct the research study “Review of hospital based Ayurvedic Electronic Health Records to gain real world knowledge - a retrospective data analysis” using I-AIM anonymized Electronic Health Records as per the protocol approved by the IEC.

I am assured that Mr. Vinay Mahajan will maintain confidentiality of the data.

Further, it is also agreed that any presentation and publication of the results arising from the study will be done after due permission from the authorities of IAIM and TDU.

Dr. Prasan Shankar
Medical director
IAIM
Bangalore

****

5 References

[1] M. Masoud Y. Jaradat A. Manasrah Ahmad I. Jannoud, "Sensors of Smart Devices in the Internet of Everything (IoE) Era: Big Opportunities and Massive Doubts," *Journal of Sensors,* p. 26, 2019.

[2] W. R. Bolislis F. Myriam T. C. Kühler, "Use of Real-world Data for New Drug Applications and Line Extensions," *Clinical Therapeutics,* vol. 42, no. 5, pp. 926-938, 2020.

[3] AYUSH Minsitry Government of India, "AYUSH website," [Online]. Available: http://dashboard.ayush.gov.in/.

[4] R. H. Singh, "Exploring issues in the development of Ayurvedic research methodology," *Journal of Ayurveda and integrative medicine,* pp. 91-95, 2010.

[5] I. Krakau, "The importance of practice-based evidence," *Scandinavian Journal of Primary Health Care,* pp. 130-131, 2000.

[6] B. Patwardhan, "Envisioning AYUSH: Historic Opportunity for Innovation and Revitalization," *Journal of Ayurveda and integrative medicine,* pp. 67-70, 2014.

[7] G. Tillu TDU hospital staff, "Exploration of the TDU database," 2016 - 2018. [Online].

[8] GOVERNMENT OF INDIA MINISTRY OF HEALTH AND FAMILY WELFARE, "THE DRUGS AND COSMETICS ACT, 1940," [Online]. Available: http://naco.gov.in/sites/default/files/Drug%20%26%20Cosmetic%20Act%201940_1.pdf.

[9] "Centre for development of advanced computing," C-DAC Pune, AAI Group, [Online]. Available: https://www.cdac.in/index.aspx?id=hi_dss_ayusoft_n.

[10] Council of Scientific & Industrial Research (CSIR), "Traditional Knowledge Digital Library," [Online]. Available: http://www.tkdl.res.in/tkdl/langdefault/common/Home.asp?GL=Eng.

[11] AYUSH ministry Government of India, "AYUSH Dashboard," [Online]. Available: http://dashboard.ayush.gov.in/.

[12] A.B. Vaidya, "Reverse pharmacological correlates of ayurvedic drug actions," *Indian J Pharmacol,* vol. 38, p. 311–315, 2006.

[13] B. Patwardhan, "Bridging Ayurveda with evidence-based scientific approaches in medicine," *EPMA Journal,* vol. 5, 2014.

[14] B. Patwardhan, "Envisioning AYUSH: Historic Opportunity for Innovation and Revitalization," *Journal of Ayurveda and integrative medicine,* vol. 5, no. 2, pp. 67-70, 2014.

[15] S. P. Kulkarni, "HURDLES IN RESEARCH IN AYURVEDA AND THEIR POSSIBLE SOLUTIONS," *International Ayurvedic Medical Journal,* vol. 3, no. 1, 2015.

[16] S. M. Baghel, "Need of new research methodology for Ayurveda," *Ayu,* vol. 32, no. 1, pp. 3-4, 2011.

[17] D. Shankar, "Health sector reforms for 21(st) century healthcare," *Journal of Ayurveda and integrative medicine,* pp. 4-9, 2015.

[18] D. Shankar, "Directions for revitalization of Ayurveda in the 21st century," *Journal of Ayurveda and integrative medicine,* vol. 9, no. 4, pp. 245-247, 2018.

[19] H. Walach T. Falkenberg V. Fønnebø et al., "Circular instead of hierarchical: methodological principles for the evaluation of complex interventions," *BMC Med Res Methodol,* vol. 29, no. 6, 2006.

[20] R. L. Byyny, Spring 2012. [Online]. Available: https://alphaomegaalpha.org/pharos/PDFs/2012-2-Editorial.pdf.

[21] A. Patrizio, "IDC: Expect 175 zettabytes of data worldwide by 2025," 3rd Dec 2018. [Online]. Available: https://www.networkworld.com/article/3325397/idc-expect-175-zettabytes-of-data-worldwide-by-2025.html.

[22] T. Cosgrove, "Consult QD: Dealing with Healthcare's Data Explosion by," [Online]. Available: https://www.google.co.in/amp/s/consultqd.clevelandclinic.org/dealing-healthcares-data-explosion/amp/.

[23] K. Adane M. Gizachew S. Kendie, "The role of medical data in efficient patient care delivery: a review," *Risk management and healthcare policy,* vol. 12, pp. 67-73, 2019.

[24] S. Kandel J. Heer C. Plaisant et al., "Research directions in data wrangling: Visualizations and transformations for usable and credible data," *Information Visualization,* 2011.

[25] "The Importance of Data Preparation for Business Analytics," 16th July 2019. [Online]. Available: https://www.ironsidegroup.com/2019/07/16/data-preparation-business-analytics/.

[26] "eTutorials.org," [Online]. Available: http://etutorials.org/Misc/data+quality/Part+I+Understanding+Data+Accuracy/.

[27] D. Tobin, "Data Transformation: Explained," 1st June 2020. [Online]. Available: https://www.xplenty.com/blog/data-transformation-explained/.

[28] Committee on Strategies for Responsible Sharing of Clinical Trial Data, "Board on Health Sciences Policy Institute of Medicine. Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk," in *The Clinical Trial Life Cycle and When to Share Data*, Washington (DC), National Academies Press (US), 20th Apr 2015.

[29] A. Kapoor, "Quality Medical Research and Publications in India: Time to Introspect," *International journal of applied & basic medical research,* vol. 9, no. 2, pp. 67-68, 2019.

[30] J. Tauberer, "Analyzable Data in Open Formats (Principles 5 and 7)," in *Open Government Data: The Book*, 2014.

[31] K. Tarsi T. Tuff, "Introduction to Population Demographics," *Nature Education Knowledge,* vol. 3, no. 11, p. 3, 2012.

[32] E. P. Balogh B. T. Miller J. R. Ball, "Committee on Diagnostic Error in Health Care, Board on Health Care Services, Institute of Medicine, The National Academies of Sciences, Engineering, and Medicine," in *Improving Diagnosis in Health Care - The Diagnostic Process*, Washington DC, National Academies Press (US), 29th Dec 2015.

[33] "PostgreSQL: The World's Most Advanced Open Source Relational Database," The PostgreSQL Global Development Group, [Online]. Available: https://www.postgresql.org/.

[34] The R Core Team, "The Comprehensive R Archive Network," [Online]. Available: https://cran.r-project.org/.

[35] "Python," Python Software Foundation, [Online]. Available: https://www.python.org/.

[36] "Java," Oracle, [Online]. Available: https://www.java.com/en/.

[37] B. Mike, "D3 Data-Driven-Documents," [Online]. Available: https://d3js.org/.

[38] "Tableau," [Online]. Available: https://www.tableau.com/.

[39] M. G. Larson, "Descriptive Statistics and Graphical Displays," *Circulation,* vol. 114, no. 1, pp. 76-81, 2006.

[40] D. Murray, Tableau Your Data! Fast and Easy Visual Analysis with Tableau Software, 1st ed., Wiley Publishing, 2013.

[41] R. McGill J. W. Tukey W. A. Larsen, "Variations of box plots," *The American Statistician,* vol. 32, no. 1, pp. 12-16, 1978.

[42] E. Tufte, The Visual Display of Quantitative Information, 2nd ed., Cheshire, Connecticut: Graphics Press, 2001.

[43] D. M. Morales-Silva C. S. McPherson G. Pineda-Villavicencio et al., "Using radar plots for performance benchmarking at patient and hospital levels using an Australian orthopaedics dataset," *Health Informatics Journal,* pp. 2119-2137, September 2020.

[44] B. McPherson, "The code to generate dynamic bubble plot from Github," [Online]. Available: https://gist.github.com/brucemcpherson/4684498.

[45] C. DeMartini, "datablick," [Online]. Available: https://www.datablick.com/blog/2018/3/14/layering-data-for-custom-tableau-visualizations.

[46] B. McPherson, "Desktop libeartion," [Online]. Available: https://ramblings.mcpher.com/.

[47] H. Hofmann, "Mosaic Plots and Their Variants," in *Handbook of Data Visualization*, Berlin, Heidelberg, Springer Handbooks Comp.Statistics, 2008.

[48] R. Wicklin, "SAS Blogs," 23 May 2018. [Online]. Available: https://blogs.sas.com/content/iml/2018/05/23/butterfly-plot.html.

[49] "Clinical Data Interchange Standards Consortium: CDISC," [Online]. Available: https://www.cdisc.org/.

[50] "TransCelerate BioPharma Inc.," [Online]. Available: https://www.transceleratebiopharmainc.com/.

[51] "U.S. Food & Drug administration," [Online]. Available: https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence.

[52] "U.S. Food & Drug administration: Framework Real World Evidence Program," [Online]. Available: https://www.fda.gov/media/120060/download.

[53] "European Medicines Agency," [Online]. Available: https://www.ema.europa.eu/en/about-us/how-we-work/big-data/data-analysis-real-world-interrogation-network-darwin-eu.