Analysis of hospital based ayurvedic clinical practice to gain real world data knowledge

Vinay Mahajan

PhD thesis

Guides: Dr. Ashwini Godbole,

Dr. Girish Tillu,

Dr. Ashwini Mathur

Contents

[Introduction 3](#_Toc41668474)

[Origin of word doctor 3](#_Toc41668475)

[Origins of pharmaceutical industry 4](#_Toc41668476)

[Elaboration of clinical trials: Origin of RCT blinded trials 4](#_Toc41668477)

[Role of statistics, analyst, programmer 6](#_Toc41668478)

[Modern hospitals, every day clinical practice and healthcare environment 7](#_Toc41668479)

[Practice of Ayurveda, reverse pharmacology, experts’ view 8](#_Toc41668480)

[Dr. Ashok Vaidya 8](#_Toc41668481)

[Prof R H Singh 9](#_Toc41668482)

[Dr. Baghel 9](#_Toc41668483)

[Dr. Ram Manohar 9](#_Toc41668484)

[Prof Bhushan Patwardhan 9](#_Toc41668485)

[Prof. Darshan Shankar 10](#_Toc41668486)

[Observational studies, Real world evidence, level of evidence 10](#_Toc41668487)

[Efforts at national level 11](#_Toc41668488)

[AYUSH 11](#_Toc41668489)

[Digital India movement 12](#_Toc41668490)

[National AYUSH Morbidities and Standardized Terminologies E- Portal (NAMASTE Portal) 13](#_Toc41668491)

[AYUSH Research Portal 13](#_Toc41668492)

[AYUSH Hospital Management System 13](#_Toc41668493)

[Electronic Medical Record, medical IT infrastructure 13](#_Toc41668494)

[Need of this study and methodological framework 14](#_Toc41668495)

[The proposed study in brief 17](#_Toc41668496)

[Chapter outline 18](#_Toc41668497)

[Converting real life clinical data into analyzable format 23](#_Toc41668498)

[Background 23](#_Toc41668499)

[Material and Methods: 24](#_Toc41668500)

[Challenges 30](#_Toc41668501)

## Introduction

## Origin of word doctor

The word doctor is a Latin word and later a French one, meaning anyone who is a teacher - usually of law, theology, philosophy, as well as medicine for a learned profession. Physician is a more accurate way to call a medical doctor, which comes from the Latin word physicum and later the French word physic, coined in 1212. Around 1400 AD, the third edition of the Shorter Oxford English Dictionary was the first to use the word “physician”. It is remarkable to note that, as per this edition, physicians and surgeons were members of entirely distinct professions, and were thought of as rivals.  
  
The work done by Szasz and Hollender (1956) defined three basic models of the doctor-patient relationship. These are (a) active-passivity, (b) guidance-co-operation and (c) mutual participation. The activity-passivity and guidance-co-operation models are completely paternalistic and largely doctor centered. The final, mutual participation has a bigger importance on patient centered treatment.  
  
Early age doctor patient relation

In older Egyptian medicine, the activity-passivity type relationship existed and remained unaltered throughout. The Greeks depended more on real-life observation, boosted by practical trial and error experience, discarding magical and religious explanations of human bodily dysfunction. In the medieval Europe, the doctor, packed with magical powers, was in a high-ranking position in society and his patients were regarded as powerless subordinates, similar to the activity-passivity model.  
  
During the early 18th Century, doctors were few in number and their patients mainly upper class and noble. This status inequality confirmed the dominance of the patient and doctors had to compete with each other in order to satisfy the patient. During the late 18th Century, hospitals came up as places to treat patients who were deprived. Doctors found themselves providing medical treatment for those who were traditionally regarded as more passive, i.e. an activity-passivity (paternalistic) model. This hospital model involved the checkup of the patient's body, the expert clinical and anatomical knowledge possessed by the doctor to make a diagnosis, making the patient dependent as a result. The association was between a dominant doctor and a passive patient, i.e. an activity-passivity (paternalistic) model.  
  
Balint (1964) maintained that “the most potent therapeutic tool the doctor possessed was himself or herself”. However, Balint recognized that very little was known about the ‘pharmacological’ attributes of this drug, such as the right ‘dosages’ (frequency of visits), any addictive properties (whereby the patient becomes gradually reliant on the doctor), and side effects (i.e. what harm the doctor could do), the doctor-patient relationship was a joint investment which over time would benefit both parties.  
  
The doctor-patient relationship in the two oldest civilizations, those of India and China, has remained far more constant than in Western societies. A patriarchal style still dominates, and doctors have a high status in society. In the ancient India, doctors were enjoying the highest level of respect in the society because of the outlook and interest towards the patient. They were deemed as “Demi Gods”. In that era medicine tradition was seen from social and philanthropic standpoint rather than taking it as a business standpoint. Through the 18th century, general physicians mostly practiced medicine in their own village or nearby area. They established a respectable relationship with the community, as they were always available. They commonly discussed with the patients regarding diverse topics of family rather than usual doctor - patient interaction. They knew the health status of all family members of the patient. As a result, a good interpersonal connection was developed between doctor and patient. Patient had utmost faith that doctor could do any harm. At times, any undesirable condition aroused then patient thought that it was only God’s will.

The vaidyas and the hakims from the ayurvedic and Unani systems of medicine complemented and used skills and concepts from each other as well as from the Arabs and Chinese. There is almost no report of enmity between them. However, the relations between the doctors of indigenous and Western medicine was anything but honest. There was a clash of cultures, the East was perceived as weak vs. the powerful knowledge of the West. In the East, medicine was mostly pluralistic. There was recognition and acceptance of alternative traditions. Medicine was not regarded only as a biological phenomenon and importance was given to a patient's societal standing, environment, and relation with the therapist. As Imperial rule intensified, declarations of the Western scientific superiority increased. Allopathic practitioners saw themselves as modernizers and often treated their indigenous peers with disrespect for their “inferior knowledge”. Local knowledge was labeled unscientific or irrational. While Western medicine was given the status of official medicine, the state turned biased and hostile toward the other systems [17]. The rising nationalism also posed to be an obstacle in a beneficial give and take of ideas [18].

Origins of pharmaceutical industry  
The roots of the pharmaceutical industry lie back with the apothecaries and pharmacies that offered traditional remedies as far back as the middle ages, offering a hit-and-miss range of treatments based on centuries of folk knowledge. The modern medicines regulation began only after revolutionary development in the 19th century life sciences, in chemistry, physiology and pharmacology, which put a robust foundation for the modern drug research and development.  
  
Unfortunate events like deaths due to diethylene glycol poisoning in the US in 1930s, the thalidomide disaster in late 1950s, have catalyzed the development of medicines regulation more than the evolution of a knowledge base throughout the world. The formation ICH consortium is one such a byproduct. The use of statistics to support R&D of new medicines has grown multifold since the Kefauver-Harris Amendments (1962). These clearly stated that the Food and Drug Administration (FDA) would require “substantial evidence” of the impact of a drug in a clinical trial setting and that new drug approvals would not be based only on proof of safety. In the USA and all over world, since 1970s, the value of medicine has been clearly exhibited by a longer life expectancy, a lower infant mortality rate, and the higher quality of life many of our senior citizens have been enjoying.

## Elaboration of clinical trials: Origin of RCT blinded trials

Randomized Controlled Trial (RCT) is a classical research design, in which the participants are randomly allocated to one or other treatment conditions under the study. RCTs help in reliable description of causality. Randomization is the fundamental characteristic of an RCT and it describes the random distribution of subjects to the study arms. RCTs provide the researcher the promise that the difference in the outcomes among subjects in study arms was exclusively caused by the intervention, as randomization equalizes the study group in all other factors. Thus, RCTs set the standard of excellence in health sciences research.

At the beginning of the study, randomization reduces the bias in assigning subjects to the intervention and control group. However, it does not rule out the possibilities of differential treatment of groups or biased adjudication of outcome variables. Blinding assists in monitoring several types of biases that might unintentionally creep into the study. The two major biases, namely performance bias and the ascertainment bias that can be controlled using blinding. The four groups of people blinded in the trial are the study subjects, the investigator(s), the outcome assessor(s), and the data analyst(s). Based on the number of people blinded, trials are classified as open label trial, single blinded trial, double blinded trial, triple blinded trial, and quadruple blinded trial.

<http://ejournal.manipal.edu/mjnhs/docs/Volume%203_Issue%201/Full%20text/9.%20Vishnu.pdf>

Physicians and clinical researchers have remained confident that RCTs deliver the most rigorous test of preventive, diagnostic, and therapeutic interventions. They are universally denoted as the “gold standard” of experimental medical analysis, as a undisputable starting point in diagnostic or therapeutic evaluation. When did RCTs become the “gold standard”? The first occurrence the authors of the referred paper have found of the phrase “gold standard” to refer to RCTs came in the pages of The New England Journal of Medicine (NEJM) in December 1982, in an article written by Alvan Feinstein and Ralph Horwitz. This date surprised the authors as a very late date for the first usage. Despite extensive searching, they have found no earlier occurrence of “gold standard” in reference to RCTs. They quote “We are eager to be proven wrong, but until all textbooks, conference proceedings, journals, and archival collections have been digitized and made full-text searchable, the gold standard of historical research itself remains elusive”.

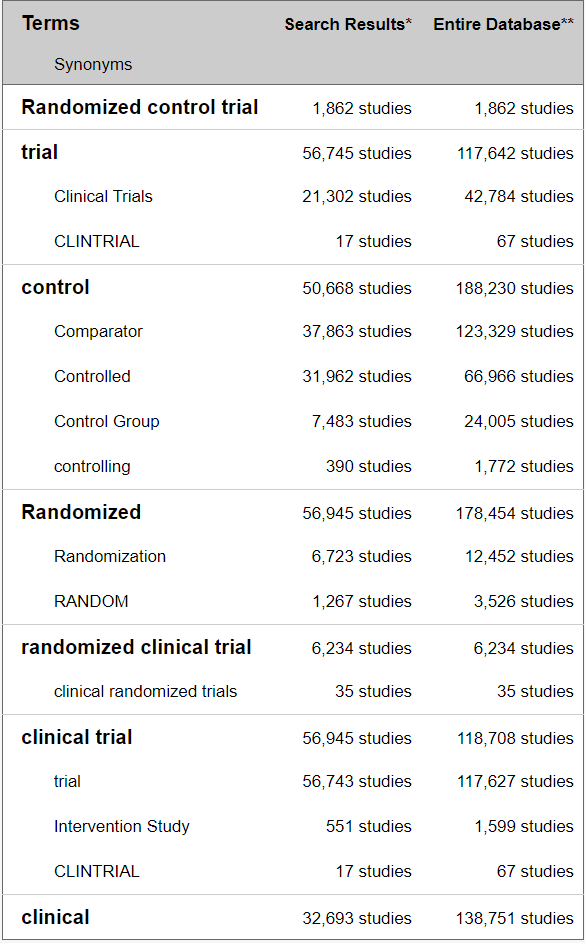
<https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(15)60742-5/fulltext>

While probing on the details regarding RCTs on the ClinTrial.gov website the following data was discovered on 24th April 2020:

* Total trials in the ClinTrial.gov database = 336905
* The estimated % of RCTs based on the results range between 2.40% (only considering “Randomized control trial” and “randomized clinical trial” 1862 + 6234) to 19.31% (considering “Randomized control trial”, “Randomized” and “randomized clinical trial” 1862 + 56945 + 6234), remaining 80% to 97% of the registered trials are not RCTs.

Screenshot from the Search Details: “Randomized control trial OR randomized clinical trial”.

<https://clinicaltrials.gov/ct2/results/details?cond=&term=Randomized+control+trial+OR+randomized+clinical+trial&cntry=&state=&city=&dist=&Search=Search>



As of April 2020, Clintrial.gov database had 19% RCTs of all the studies registered in the database. This provides us with enough evidence that a lot of research work is carried outside of the RCT framework.

## Role of statistics, analyst, programmer

The practice of biostatistics, or statistics in biomedical research, has developed by quantitative methods from other areas such as mathematics, operations research, and economics. Statistics is regularly applied to genomics, population genetics, epidemiology, ecology, nutrition, etc. The 1990s presented explosive increase in applications of computationally demanding methods originated on statistical principles, due to the focus on translational research, personalized medicine, and the relevance of real-world data. Statistical modeling has become more complex due to the size of data, computational requirements and varied data sources. These developments presented additional roles like statistical programmer to statistical analyst to clinical programmer to clinical analyst.  
  
A clinical data analyst (or clinical informatics analyst) is a healthcare professional responsible for confirming the validity of scientific experiments and data gathered. They apply their knowledge of data acquisition, management, analysis, and interpretation to healthcare data, providing actionable insights that physicians, clinical researchers, decision-makers, and others can use. They may be responsible for automating internal and external reports, creating executive-level dashboards, and presenting information to help hospital executives and others understand the operational impact of the data. These data analysts ensure that processes and protocols are followed, thereby improving the quality and efficiency of care.

They help organizations improve the quality of care, lower the cost of care, and enhance the patient experience. They provide data insights that drive clinical process improvement, such as reducing readmissions and hospital-acquired conditions. In addition, healthcare analysts help insurers, vendors, and others synthesize data that guides decision-making, population health management, cost containment, and quality improvement.

Clinical data analysts report the results of clinical business intelligence to management, stakeholders, and other interested parties. They coordinate with other relevant departments such (e.g., clinical strategy, clinical operations) to determine the areas to be analyzed as well as the appropriate measures that should be taken to ensure data analysis proves useful.  
  
This role has the ability to combine strengths from multiple areas to build a powerful storyline. By providing greater insight to patients, providers, and policy makers into the appropriate application of interventions, and quality and costs of care, these data offer the opportunity to accelerate progress on the six dimensions of quality care—safe, effective, patient centered, timely, efficient, and equitable (Chaudhry, 2006; IOM, 2001, 2009; Safran et al., 2007).

## Modern hospitals, every day clinical practice and healthcare environment

One trend all modern hospitals have in common is that the amount of processed information generated per patient is constantly increasing EMR, patient charts, CT scans, X-rays, ECGs, pathology reports, wearable devices, sensor data, etc. When such information is properly compiled and aggregated, it provides data for efficient hospital administration. How each patient is treated can yield important statistical data, while the number of patients handled, grouped by sex, age, and diagnosis, provides basic information required by administrators.  
  
Health information provided by the hospital is vital for public health administrators at the district level. Such data, accumulated from all the district's hospitals, are essential in formulating health planning for the district, by matching with population statistics, other demographic data and the health resources of the district. This medical information, containing not only data about the patient's illness but also the success or failure of therapy, is a precious source of clinical training for medical students, nurses and related health care work force. This is a basis of new medicine and therapies. The development of new therapeutic drugs is built upon the careful observations of experienced physicians, and the comparison and analyses of the effects of previously administered drugs, gathered from the patient's data. This in turn leads to continued medical progress in, say, uncovering a new entity of a disease or a new method of diagnosis. This information is processed manually or partially automatically. This inefficient method results in delays in reaching crucial treatment decisions, for instance when test results fail to get back to the physicians in time, or in delayed action by administrations when a prompt response is needed.  
  
Internet or google era  
In the last decade of the 20th century and starting of 21st century, a chronic displeasure began to settle over the relationships between doctors and patients. Due to multiple advances in medical, chemical, bio-chemical, IT industries, patients were doing better but feeling worse, the personal, solid and faithful relationship is failing.

Earlier, patients were most often considered medically illiterate to make decisions on their own behalf. Thus, informing patients about the uncertainties and shortcomings of medical interventions worked only to damage the faith that was so important to the therapeutic success. Doctors felt secure in making decisions on behalf of their patients. Later on, the distance between the doctor and patient widened. Little social mixing continued, and the doctor-patient relationship became impersonal and isolated, based upon negotiation and financial transaction.

Modern-day effect on the doctor-patient association has been the exponential growth in the use of the Internet by patients. This has meant that patients are largely better informed. In an effort to understand more about the cause of their symptoms through Internet searching, patients assume that they are suffering from a serious disease, which is often not the case, say doctors. They have warned that self-diagnosis through Internet might be misleading for patients.

## Practice of Ayurveda, reverse pharmacology, experts’ view

The western medicines are developed using a method called as hierarchical method where it tries answering the 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, Open-label randomized studies. The process ends with the blinded, randomized, placebo controlled trials (RCT). The RCTs offer most internal validity and reduce the bias. These studies could be complemented by then moving onto case studies, case series. This “one step at a time” approach has worked very well in the western medicine framework.

There are some other 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, huge observational data for ayurvedic medicines covers: more than 1,00,000 books and manuscripts, 57 authentic books (Drug and cosmetic act 1940), > 4500 diseases including subtypes and conditions (Ayusoft database), > 81,000 formulations (TKDL database), > 4,00,000 Practitioners (Planning Commission - 11th Plan) in India, infinite documents, references, experiential data, living tradition and knowledge in public domain. Dravyaguna (Pharmacology), Bhaisajya Kalpana (Pharmaceutics), Nidana (Diagnosis) and Chikitsa (Management principles). This data points to a validated knowledge base and it is acceptable that Ayurveda is an evidence based knowledge system. Let us take a look at points of views of some of the leading figures in the field:

### Dr. Ashok Vaidya

Dr. D. B. Vaidya has explained the concept of reverse pharmacology to understand the action mechanism of Ayurvedic intervention. Reverse pharmacology is the science of integrating documented clinical/experiential hits, into leads by trans-disciplinary exploratory studies and further developing these into drug candidates by experimental and clinical research. It comprises of three stages - experiential, exploratory and experimental.

1. Experiential robust documentation of clinical observations of the biodynamic effects of standardized ayurvedic drugs by meticulous record keeping.
2. Exploratory studies for tolerability, drug interactions, dose range finding in ambulant patients of defined subsets of the disease and para-clinical studies in relevant in vitro and in vivo models to evaluate the target activity.
3. Experimental studies, basic and clinical, at several levels of biological organization, to identify and validate the reverse pharmacological correlate of ayurvedic drug safety and efficacy.

Based on the huge observational data and the relatively low rate of side effects, it is easy to test Ayurvedic intervention in larger clinical trials. This would help build the required safety and efficacy information relevant to the Ayurvedic intervention under question. Once these key parameters are established, the pharmacokinetic properties can be understood. This process is economical and may take lesser amount of time when compared with the hierarchical model used in western medicine.

### Prof R H Singh

Research experiments within Ayurvedic area undertaken during the last 50 years have not been very rewarding, except for the extremely useful exercise of literary research. These have at least made a few of the classical Ayurvedic texts accessible to contemporary readers and researchers. A number of literary reviews published in recent years have helped create a conceptual interface between Ayurveda and modern science. However real laboratory-based new research is still awaited. Such a scientific deadlock deserves exploiting newer strategies, while keeping with the fundamental principles of Ayurveda-as-it-is, without distorting it to suit the application of modern research technology. In any research, the goal of research should not be compromised to suit the convenience of research methods. But unfortunately in Ayurvedic research, there has always been a reverse compromise, and in my perception this attitude is the main reason for failures in this otherwise potentially most fruitful field of contemporary medical research. The so-called scientific research of several decades has helped neither Ayurveda nor modern medicine to any significant extent except in creating awareness.

<http://europepmc.org/article/PMC/3151394>

### Dr. Baghel

Research should be a process that converts data into information, information into knowledge and knowledge into wisdom. It should be balanced, comprehensive, and equally emphasizing in the literary field, experimental and clinical research. It should be able to impact the fields of academics, pharmacy and practice in a profound way. The researches done in the last 60 years on Herbal Pharmacology have led confirmation of few concepts like Reverse Pharmacology and use of whole crude drugs in place of isolation of fractions for clinical trials. Various researchers started to feel that conventional clinical trial regimen is not fit for Ayurveda. Ayurveda is a pure science based on strict logical explanation, which is called *Darshana*. Ayurveda was always in the developmental phase like all the medical systems should be.

### Dr. Ram Manohar

He has opined that Ayurveda is based on 5000 years of clinical practice. Hence, in place of conventional evidence-based medicine (EBM) clinical trials, practice-based clinical trials should be organized for Ayurveda.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3215413/>

### Prof Bhushan Patwardhan

The holistic concepts of Ayurveda give emphasis to health promotion, disease prevention, early diagnosis and personalized treatment. There seem to be substantial similarities between the traditional systems like Ayurveda and the innovative approach of predictive, preventive and personalized medicine (PPPM). Continuous research on safety, quality and efficacy of Ayurvedic drugs and procedures is needed. Systematic documentation and critical analysis of clinical practice are necessary. It is very important to review available evidence in the right perspective. In case of Ayurveda, the evidence can be drawn from two main sources.

1. Source of evidence may be based on historical, classical and present nature of clinical practice. Here, the documentation of practice to support various claims is very crucial. Mere reference to classical texts is not sufficient as evidence for practice.
2. Source of evidence may be based on scientific research to support various theories, medicines and procedures used in Ayurvedic medicine. A critical situation analysis of present status of clinical practice and scientific research on Ayurvedic medicine may be necessary at this stage.

In another paper, he remarks “We need contemporary practices and models for healthcare that bring confluence of Ayurveda and modern science. Becoming modern is not a crime; it does not prevent us from maintaining our own cultural identity. For instance, Charaka would not have ignored technologies like electron microscope if they had been available during his time”.

<https://link.springer.com/article/10.1186/1878-5085-5-19>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4061590/>

### Prof. Darshan Shankar

As of 2018, at the national level, Ayurveda receives a small outlay of around 2.5% of the Central health Budget and at the State level, it receives a minuscule part of the state health budgets. Despite its strengths, Ayurveda definitely has some limitations in current scenario. Today, Ayurveda needs to interface in an epistemologically informed manner, with molecular biology in order to discover its own mode of actions at the structural level and embrace tools of information technology to organize its enormous multifaceted data, in searchable formats. Rigorously documented clinical experiences interpreted through Ayurveda-biology will deepen modernization of healthcare in India.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6314269/>

## Observational studies, Real world evidence, level of evidence

Real-world data are data captured in an observational manner, in a natural, uncontrolled setting – outside of traditional clinical trials. Ordinary clinical practice, producing a never-ending flow of results from every day practice, can be viewed as infinite sequence of unsystematic observational studies. There must be a rational use of outcome data from patients, far more representative of the general population than those included in formal clinical trials. Practice-based medicine is an important way to advance science. For this to happen, the treating doctor should be a “clinician researcher investigator”. Doctors who, even in their busy practice, are able to see patterns in their practice, then formulate a research hypothesis, and then proceed to test it. Thus, a doctor can be good at both good clinical practice (GCP) and good clinical research practice (GCRP). One can feed the other and vice versa, and this way medicine advances.

Investigation of the literature across time points from year 2000 to year 2020 presents that well designed observational studies could be producing treatment effects similar to the RCTs:

**Year 2000:** For the five clinical topics and 99 reports evaluated, the average results of the observational studies were remarkably similar to those of the randomized, controlled trials. The results of well-designed observational studies do not systematically overestimate the magnitude of the effects of treatment as compared with those in randomized, controlled trials on the same topic.

John Concato, Randomized, Controlled Trials, Observational Studies, and the Hierarchy of Research Designs, N Engl J Med 2000;342:1887-92

**Year 2007:** “When done correctly, epidemiologic studies of drug effects can be both more conceptually demanding and more powerful than the average RCT, especially in assessing drug safety. RCTs offer one kind of knowledge but prevent us from seeing other properties of a drug. Epidemiologic studies can help elucidate those properties but may introduce new blind spots. We must do both kinds of research with rigor and with humility.”

Jerry Avorn, In Defense of Pharmacoepidemiology — Embracing the Yin and Yang of Drug Research, N Engl J Med.2007:357;22

**Year 2014:** Our results across all pooled reviews (pooled ROR: 1.08) are very similar to the results reported by similarly conducted reviews. As such, we have reached similar conclusions, on average, there is little evidence for significant effect estimate differences between observational studies and RCTs regardless of specific observational study design, heterogeneity, or inclusion of studies of pharmacological interventions. Factors other than study design per se need to be considered when exploring reasons for a lack of agreement between results of RCTs and observational studies. Our results underscore that it is important for review authors to consider not only study design, but the level of heterogeneity in meta-analysis of RCTs or observational studies. A better understanding of how these factors influence study effects might yield estimates reflective of true effectiveness.

Plain language summary: healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials (review) 2014, The Cochrane Collaboration.

**Year 2020:** Accumulating evidence suggests that appropriately conducted RWD studies have the potential to support regulatory decisions in the absence of RCT data. Further work may be needed to better illustrate the settings in which RWD analyses can robustly and consistently match the results of RCTs and, more importantly, the settings in which they cannot match them. After careful consideration of the potential for bias, regulators can then determine when they would unequivocally accept RWD in place of an RCT. If studies based on RWD are ever to replace RCTs, regulators may need to accept that the cost of accelerating patient access to treatment carries a higher level of decision-making uncertainty than that with which they are familiar.

Ramagopalan, S.V., Simpson, A. & Sammon, C. Can real-world data really replace randomised clinical trials?. BMC Med 18, 13 (2020). https://doi.org/10.1186/s12916-019-1481-8

Practice uses knowledge, focuses on individual patient, has a short action span, the reward is immediate, it regards authority, follows custom, earns income, and encourages research. On the other hand, Research creates knowledge, focuses on groups, has a long action span, the reward is delayed, it questions authority, challenges custom, earns reputation, and enriches practice. There have been many attempts to generate clinical evidence from primary health care by systematic utilization of patient records. Many examples have been cited for the (1) deficient clinical data, (2) in-accurate input giving "garbage in garbage out", (3) insufficient follow-up, (4) very few fully completed case records with risk factors, co-morbidities, etc. With standardized protocols, checklists and assessment scales constituting the hard core of a renewed medical record in primary care, good clinical practice would be better supported and even integrated with research on a regular basis. That is how future clinical research, including observational studies and experimental trials, will be based on the practice and principles of primary care.

## Efforts at national level

### AYUSH

The Ministry of AYUSH is an initiative by the Government of India to promote the propagation and development of AYUSH systems of health care and medicine in India (Ayurveda, Yoga & Naturopathy, Unani, Siddha and Homoeopathy). To administer the different medicine systems encompassed by the Ministry of AYUSH, it has five research councils or departments, affiliated courses, and affiliated national institutes.

### Digital India movement

The Digital India Movement initiated by Govt. of India has made significant changes in IT sector. We are witnessing tremendous improvement in application of Information Technology (IT) in every Ministry and its organizations. Health is a major concern in India. Implementation of IT in healthcare is being accelerated to provide effective and better care for citizens of India.

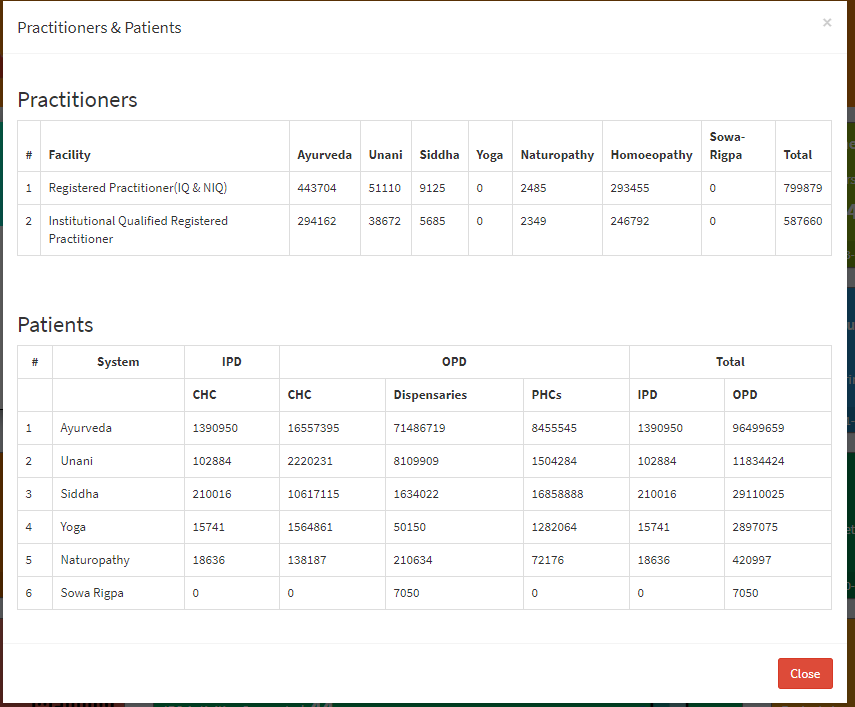
Ministry of AYUSH has taken initiative to get onboard in healthcare Information technology implementation projects. In this regard, Ministry of AYUSH has created AYUSH GRID to get all the IT projects under one umbrella. It is integration of IT projects exclusively meant for improvement and facilitation of AYUSH pan India.

The initiative is a poised to emerge as a game changer as it is the first of its kind citizen centric service from Government of India to provide Electronic Health Record and Personal Health Records facility to the recipients of AYUSH Health Services catered through Research Councils, National Institutes and other related agencies under Ministry of AYUSH, Government of India.

The dashboard available on AYUSH homepage (accessed on 25th May 2020), reveals great facts on various components.



Details of the number of ayurvedic practitioners and total number of patients treated in the picture below show almost 10 crore patients treated at some or the other point. There are approximately 140+ countries with less than 10 crore population, which provides a perspective on the size of data available at the AYUSH level. It remains to be seen how to convert “data into information, information into knowledge and knowledge into wisdom”.



### National AYUSH Morbidities and Standardized Terminologies E- Portal (NAMASTE Portal)

CCRAS along with other stakeholders has developed the portal. It provides information about Standardized terminologies and Morbidity codes along with dedicated data entry module for updating morbidity statistics in consolidated form as well as on real time basis.

<http://namstp.ayush.gov.in/#/index>

### AYUSH Research Portal

To disseminate the merits of AYUSH systems across the globe, a web-based portal for Research publications in AYUSH was launched in 2011, which is being maintained by NIIMH Hyderabad. The portal is successfully continuing and the information is being updated periodically.

<http://ayushportal.nic.in/>

### AYUSH Hospital Management System

A-HMIS is a comprehensive IT platform to effectively manage all functions of health care delivery systems and patient care in AYUSH facilities. THERAN (THE Research Application Nexus) - Hospital Management Information System developed by Siddha Central Research Institute, Chennai under Central Council for Research in Siddha, Ministry of AYUSH was near to the requirement to give complete solution of HMIS amongst various HMIS present in AYUSH Institutions of Govt. Of India, hence it was decided to upscale and customize it as per systems in AYUSH.

## Electronic Medical Record, medical IT infrastructure

In early 2000s countries like Canada, UK, New Zealand, Estonia, etc. started collecting data at national level. The key challenges experienced while implementing EMR were: (1) infrastructure creation, (2) Policy & regulations, (3) Standards & interoperability, and (4) Research, development & education.

India has a mixed system of healthcare consisting of a large number of government run hospitals, the private hospitals, family doctors and private medical practices. We see this trend reflected in the actual health seeking behaviour of communities where people tend to combine medicine systems like Allopathy, Ayurveda, Siddha, Sowa Rigpa, Unani, Homeopathy and Yoga depending on the nature of the disease [Darshan 1].

Prevention, wellness and treatment of Non communicable Diseases (NCDs) are a few of the core strengths of Ayurveda. The Ayurvedic strategies for immunity (*vyadishamta*) and tissue regeneration (*rasayana tantra*) can lead to new research. Its systemic pharmacology (*dravya guna shastra*), its thousands of brilliantly designed food and drug formulations (*bhaishaj kalpana*) with undiscovered pharmacokinetics and dynamics can feed a new paradigm of drug discovery for syndromes and personalized nutrition science (nutrigenomics) for decades to come. Rigorously documented clinical experiences interpreted through Ayurveda-biology will deepen modernization of healthcare in India [Darshan 1].

Ayurvedic vaidyas usually use paper based case report to record a patient’s ayurvedic parameters along with other details of medical consultation. These are not typically exchanged with other vaidyas. Increased use and interoperability with electronic medical records of digital Ayurvedic patient management systems is required. Based on a report published by AYUSH above, there are 4.5 lacs registered Ayurvedic practitioners. Even if 5% of doctors start using EMRs, i.e. 22500 doctors and if data for 2 new patients is entered every day (~225 working days) for the whole year, 50 lacs unique patient data can be generated in a single year. Currently, this gold mine of data is not yet built.

In general, the level of use of IT in the healthcare sector in India has been lower in comparison to other countries. The use is further lower in case of ayurvedic practice. The increased penetration of mobile phones and the availability of high-speed Internet offer the possibilities to provide healthcare services across the whole country. The Ministry of Health & Family Welfare and Department of AYUSH have taken several steps to create standards. It has issued guidelines for EMR standards and meta-standards. The implementation and penetration are yet to increase.

EMR data has been widely used for analysis and many papers have been published. These have generated supportive data for a variety of clinical outcomes, evaluation methods, and implementation of new technology or intervention along with awareness of unintended consequences; thus supporting the clinical decisions and aiding to improve the healthcare process or clinical outcomes.

## Need of this study and methodological framework

All of the thought leaders cited above point to the strengths of Ayurveda as well as the immediate needs.

* They have written about the research impact within the Ayurvedic paradigm
* 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 for new ways of tackling problems

The research must be directed at increasing patient benefit, while keeping it as cost-effectively as possible to permit efficient and viable service. These discussions give rise to the methodological framework covering short, mid and long term impacts, framework categories are as follows:

1. Policy makers – AYUSH and other relevant ministries,
2. Healthcare providers - Ayurveda Healthcare systems, General healthcare systems,
3. Hospital managements and individual clinicians,
4. Patients and
5. Universities and learning institutes – clinical communication, researchers to build vital evidence-base.

The following section defines a few questions falling in various categories covering frameworks listed above.

* Evidence base generation (EBM): EBM aims to improve quality of care through the integration of best research evidence with clinical expertise, patient and parents’ preferences.

|  |
| --- |
| 1. Is it possible to formulate an answerable clinical question (PICO format)? Patient, Intervention, Comparison, Outcome 2. Find the evidence 3. Understand the evidence 4. Apply the evidence 5. Evaluate the performance |

* Creation of a cohort and generating robust information

|  |
| --- |
| 1. Understand the natural history 2. Cohort identification 3. How to create a cohort of similar patients? What is the meaning of cohort in ayurvedic area? 4. Once the cohort has been designed then would it be possible to predict any questions, trajectory of diseases, etc. 5. Predicting the next complication or next disease, when and what? 6. Quantifying the effect of intervention 7. Construct evidence based guidelines |

* Description of the hospital database / operational insights

|  |
| --- |
| 1. How many patients are available in the database? Are there some simple summaries available summarizing the underlying data? 2. What are the characteristics of these patients? 3. What is % of males and females? 4. What is the age group? 5. Which countries, states, cities do they come from 6. How many times do they visit hospital? 7. What is the visit pattern? 8. What is the number of In Patients & Out Patients? 9. What kind of assessments are done at each visit? 10. What is the duration of visits overall for a patient? 11. What kind of diseases are reported? 12. How many diseases do they have? 13. What operational insights could be drawn? |

* Disease variations as seen in the database – does it point to epidemiology of Bengaluru?

|  |
| --- |
| 1. How many times the same disease is reported by a patient? 2. Are there differences seen in the diseases by gender? 3. Are there differences seen in the diseases by season? 4. Are there differences seen in the diseases by age group? 5. "How many disease - disease combinations can be identified,    1. some could be clinically meaningful,    2. some may be clinically meaningless    3. some could be rare combinations 6. Which diseases are reported a lot more than other diseases? 7. What kind of diseases are reported before an onset of a disease? 8. What kind of diseases are reported after an onset of a disease? 9. What is the duration of visits for each disease (episode, different episodes) for a patient? 10. How many doctors treat one patient through the course of their diseases at the hospital? 11. Which are the rare diseases identified in the database and any clinically meaningful document could be written? |

* Treatments prescribed and their variations as reported in the database

|  |
| --- |
| 1. How many different treatments are prescribed? 2. "How many different ayurvedic services are offered to these patients? 3. Which ayurvedic procedures are ordered for which diseases?" 4. What kind of treatments prescribed? 5. Any description of treatment types? -herbo mineral, classical, rasaushadhi, etc. 6. "Summary statistics for treatments    1. frequency    2. how many days a treatment is prescribed for    3. total duration of treatment prescription" 7. What treatment and treatment combination appears the most? 8. What is the total duration per treatment found, and does it provide any clinical meaning? 9. "Medicines based on metal formulations are under attack from non ayurvedic community, 10. What insights can be drawn about the rasa aushadhis?     1. which are the medicines?     2. for what diseases are they given?     3. for what duration?     4. Before providing the metal based treatment and after providing the metal based treatment, is there any difference in duration seen in treatments?     5. % of patients provided these medicines?     6. % of duration of all the duration of treatment given to these patients |

* Regulatory / insurance / education point of view

|  |
| --- |
| 1. How can anyone use these data as "secondary use"? 2. Insurance companies related aspects of data which has not been thought of so far? 3. Do the approved labels of medicines and prescriptions in the database match each other? 4. How can this framework be used as a supplementary material for any MD student? 5. Can TDU initiate registry for ayurvedic data from other hospitals, private ayurvedic doctors? 6. In absence of a true “hard clinical endpoint” what kind of “endpoints” be created?    1. short term outcomes (intent adeherence to the treatment)    2. intermediate term outcomes (actual adhere)    3. long term outcomes (overall health) |

* Research impact / operational impact

|  |
| --- |
| 1. Hospitals in India or any part of the world focus on treatment and not on research publications -- can a team be put together for the medical communication who are publishing papers as their primary job? 2. If there is no known profile of patients visiting the ayurvedic hospital, if this data is represented in the right form then will it provide novel information? 3. How can we measure the strengths and weaknesses of the practice? 4. How can we innovate and improve? 5. How to evaluate changes in results over time? 6. Is there a way to execute any longitudinal analysis? 7. Is it possible to build new hypothesis? 8. Is prescribed treatment truly personalized or are there enough overlaps? 9. Is it possible trace back the treatment regimen followed as per the classical fundamentals or not? 10. Is there a way to compare the demographics of patients from our hospital against the main stream western hospital? 11. How to corroborate information coming from     1. Patient level unit     2. Visit level unit     3. Disease level unit     4. Medical level unit |

* Information system challenges on a day-to-day basis

|  |
| --- |
| 1. How can our experience of doctors using the Health Information system be viewed?    1. do doctors like data entry?    2. do they see this as an additional burdern or an integral part of the work?    3. do they see the data entered by them reviewed at a later point by someone else "intrusion into their work |

## The proposed study in brief

Ayurveda has been practiced over many centuries in India. It will be safe to assume that the conceptual developments in ayurvedic knowledge base have taken place through every day observations and basic laws of nature. These fundamentals have been adjusted to the relevant times as per the passage of time, which is quite evident from vast literary history of Ayurveda, which covers subjects like pharmacology, principle of diagnosis and treatment for all branches of medicine and surgery, philosophical framework and logic, pharmacy and numerous pharmacopeias. Traditional texts enumerate more than described for each disease condition.

Generating credible evidence for such a large pool only through modern experimental means such as trials is very challenging. Methodologists are challenging current hierarchical evidence model and the circular model comprising observational research methods are proposed for CAM research. Ayurveda like any other system of medicine, is practiced more in clinics than in clinical research setting, 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.

The TDU IAIM Ayurvedic hospital database is a pioneer in capturing electronic health or medical records (EHR, EMR) since year 2011. Till October 2017, it contained data for >51,000 patients, >1,50,000 visits, > 900 variations of disease types, >3,000 variations of medical procedures. This study will be one attempt in plugging the gap of missing empirical evidence, systematic analysis of observational clinical data. It is focused on IAIM clinical data for study of efficacy and safety trends. Given the availability of the data, we have an opportunity to unearth wealth of knowledge. This kind of study has never been carried out within ayurvedic area, providing us with the pioneering chance.

### Chapter outline

The subsequent chapters of the thesis are defined as follows:

1. Converting clinical life data into analyzable format
2. Clinical data understanding
3. Studying demographics and patient specific factors
4. Diagnostics and treatment data
5. Outcome and effect
6. Prediction / NLP

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Objective s as Chapters | Hospital management | Hospital physicians | Patients | Policy makers | Ayurvedic Healthcare systems | General healthcare systems | Clinical communication | Research related impact | Future directions |
| 01 Converting clinical life data into analyzable format | x | x |  |  |  | x | x |  |  |
| 02 Clinical data understanding |  |  |  | x | x |  |  |  |  |
| 03 Studying demographics and patient specific factors | x |  | x |  |  |  |  |  |  |
| 04 Diagnostics and treatment data |  | x | x | x | x | x |  |  |  |
| 05 Outcome and effect |  |  |  |  |  |  | x | x |  |
| 06 Prediction / NLP |  |  |  |  |  |  |  |  | x |

**Potential opportunities within ayurveda:**

1. People are realizing the importance of Ayurveda and are turning towards it across world
2. Sustainable treatment solutions for non-communicable diseases readily available
3. Every day-to-day practice generates huge amounts of data, which is fragmented in nature; most of the data is stored on paper in individual practice:
   1. The size, quantity and depth are unknown, make people aware of the hidden potential
   2. Simple user friendly and effective data review tools for improved patient and doctor interaction
   3. Enhance the existing evidence further from day to day clinical practice data
4. Over the years, teaching methodology has not changed substantially and quite a lot of modern methodology, learning objectives of different kinds:
   1. Make complex relationships easily visually available
   2. Very few large scale studies like this have taken place in Ayurveda – explore the data and describe the findings, textually as well as diagrammatically
   3. Accumulate data for experienced vaidyas as well as new ones
5. Data collection and its usage at different units
   1. Doctor's unit of assessment: Whole patient
   2. Researcher unit of assessment: Disease condition, health condition
   3. Hospital Management Information System (HMIS) unit of documentation: Patient Visit
6. Do two practitioners follow a similar pattern of regimen in treating patients?

**Needs of transdisciplinary approaches:**

1. Create Transdisciplinary evidence to increase the scientific understanding outside of the community, then increase the confidence and thereby widening the user base
2. 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, treatments prescribed
3. Generate Real World Evidence (RWE) to supplement classical understanding which is rooted within the community
4. Make recommendations to the practitioners for standardized way of data collection, analysis and reporting which will support future RWE studies
5. Understand the hidden wealth of data for Transdisciplinary expansion of thoughts

**Why do we need transdisciplinary approaches: how strengths of multiple approaches should be used – how can multiple disciplines can cover the shortcomings – how 1 + 1 can become 5 and not 2? How to compensate for shortcomings and gain benefits.**

**Untapped potential of ayurveda (to make the communication more powerful, the communication is not global very localized, but if it has a potential of global application, this will make it wide open for a bigger audience, it is under-utilized as a public health tool, due to insufficient data, these solutions may help us bring it to the fore in public health, promotion of health, prevention of diseases):**

1. Shlokas to diagrams – generating supplementary material to classical texts
2. Numerical evidence for the shlokas (shastra tatva, vyavahar)
3. Thought provoking work to generate new needs, unconventional use of the data
4. Expand the use of modern IT solutions like IT infrastructure, electronic health records, cloud, etc. within ayurvedic area where appropriate – Ayur IT solutions.
5. Take advantage of freely available cutting edge software(s) to create new approaches
6. Introduce statistical programming as a tool to ayurvedic area
7. Create a frame work of data analysis – may be available to MD students, replace their individual studies by this retrospective work
8. Generate viable financial model for making data available for insurance

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5950611/>

<https://www.tandfonline.com/doi/pdf/10.1080/028134300453278>

<https://bestpractice.bmj.com/info/toolkit/practise-ebm/why-do-we-need-ebm/>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5299804/>  
<https://www.sas.com/storefront/aux/en/sppharmastat/60>

<https://www.sas.com/storefront/aux/en/sppharmastat/60622_excerpt.pdf>  
https://www.npr.org/2011/01/28/133306341/Science-Diction-The-Origin-<https://www.npr.org/2011/01/28/133306341/Science-Diction-The-Origin-Of-Physician>  
https://www.bartonassociates.com/history-of-physicians<https://www.bartonassociates.com/history-of-physicians>  
https://www.google.co.in/amp/s/m.hindustantimes.com/pu<https://www.google.co.in/amp/s/m.hindustantimes.com/pune-news/if-you-search-for-symptoms-online-when-sick-you-might-be-suffering-from-google-syndrome/story-lkbwf7x67gDsslS2yOfWbJ_amp.html>  
https://www.nepjol.info/index.php/AJMS/article/download/13929/12040<https://www.nepjol.info/index.php/AJMS/article/download/13929/12040>  
https://pharmaphorum.com/articles/a\_history\_of\_the\_pharmaceutical\_i<https://pharmaphorum.com/articles/a_history_of_the_pharmaceutical_industry/>  
https://www.who.int/medicines/technical\_briefing/tbs/Drug\_Regulation\_Histor<https://www.who.int/medicines/technical_briefing/tbs/Drug_Regulation_History_Present_Future.pdf>  
https://www.ncbi.nlm.nih.gov/books/NBK54296/<https://www.ncbi.nlm.nih.gov/books/NBK54296/>  
https://apps.who.int/iris/bitstream/handle/1<https://apps.who.int/iris/bitstream/handle/10665/46981/WH-1989-Aug-Sep-p6-8-eng.pdf?sequence=1&isAllowed=y>  
https://searchbusinessanalytics.techtarget.com/definition/data-exploration<https://searchbusinessanalytics.techtarget.com/definition/data-exploration>  
https://core.ac.uk/download/pdf/34578672.pdf<https://core.ac.uk/download/pdf/34578672.pdf>  
https://www.hpc2n.umu.se/sites/default/files<https://www.hpc2n.umu.se/sites/default/files/events/para06/papers/paper_213.pdf>

Diagrams needed, (flowchart):

Differences in the diagnosis (ayurvedic practice)

Modern medicine: with same disease, biomarker (surgery or medicine)

Ayurveda:

Put an emphasis on why RCT is a challenge for ayurveda.

RCTs may not be applied due to the complexity, so ayurveda suffers.

If we use the latest analytical tools it may uncover authentic information, unbiased, in absence of RCTs, using these tools, we may be able to get placebos are difficult to develop,

Do not use the reductionist approach – we would reductionist approach.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Hospital management | Hospital physicians | Patients | Policy makers |

# Converting real life clinical data into analyzable format

## Background

During the passage of time, revolutions in technology have continually increased the creation of information and its exchange. When humans advanced from spoken communication to written records they could steadily produce documents, papers, texts, books that some could read and learn from. Growth of writing were compilation of articles, tables and other records, which grew to become 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.

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 released in Dec 2018, that the cumulative sum of the world’s data will grow from 33 zettabytes this year 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 expansion of data has been seen in every industry, in every corner of the world. The unending increase in the quantity and flood of information denotes an important professional opportunity and a challenge at the same time for those in medicine and science.

As indicated by Toby Cosgrove MD, from Cleveland, medical information will soon double every 73 days, in year 2020, which used to take approximately 3.5 years in 2010. An estimated 800,000 papers are published yearly in 5,600 medical journals. It is projected that 12,000 new articles and 300 randomized controlled trials are added to Medline each week, and that new medical articles appear at a rate of one every 26 seconds. 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 of the data. Data can be produced, obtained, and stored in a numerous number of structures.

Raw data, like unrefined gold buried deep in a mine is a precious resource. It is often:

* Inconsistent. It contains both relevant and irrelevant data
* Imprecise. It contains incorrectly entered information or missing values
* Repetitive. It contains duplicate data

Before anyone can benefit from raw data, it needs to be extracted, filtered through, understood, and transformed into something could be analyzed. 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”.

Data access, extraction, cleansing, transforming, making it clean and consistent data are a few steps of data preparation or data wrangling. 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. The subsequent sections will provide us with detailed information to go from an observation to information from a database point of view.

## Material and Methods:

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 (e.g., X-ray, magnetic resonance imaging [MRI]), procedure results (e.g., electroencephalogram [EKG], endoscopy), or self-reported data (e.g., symptoms, quality of life).

I would present a flow diagram of how the hospital visit unfolds creating data through patient and doctor interaction.

|  |  |  |
| --- | --- | --- |
| Step | Description | Pictorial representation |
| 1 | A patient visits a hospital | C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\9E73FB80.tmp |
| 2 | Medical data is populated, either on paper or electronically | Electronic Health Record Stock Illustrations – 823 Electronic ... |
| 3 | Medical assessments are done if prescribed   1. Pathology reports 2. ECGs 3. CT scans 4. Ayurvedic tests 5. Panchakarma | C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\840CC8A6.tmp  Heartbeat Line Icon. Heart Rhytm. ECG. Cardiogram Stock Vector ...  C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\23EF0755.tmp  C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\DD1FC92B.tmp  C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\44C8A771.tmp |
| 4 | All the above combined together generates the Electronic health record / Electronic Medical record is generated  This can be called as “raw data” or “source data” | Clipart Electronic Health Record |
| 5 | The data from pathology labs, ECGs, CT scans come from multiple labs, hence there is a need of “data transfer” from one location to the source database | Paper Sheets Having Indicating Arrows Line Icon Data Transform ... |
| 6 | As seen in step 3: Source data can come from many sources, hence there is a need to integrate data sources into 1 source called as any one of the following:   1. source database 2. data ware-house 3. data lake 4. data mart | C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\ACE814D8.tmp  C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\E79DFCBA.tmp |
| 7 | Once the source data is electronically made available, it has to be made accessible through a system | Monitor Screen Check Mark Symbol Padlock Data Access Icon — Stock ... |
| 8 | This point onwards, the data preparation steps begin   1. Data extraction from source to an area 2. Data is transformed as per the needs 3. Data is loaded into areas for the “end users” to use per needs | C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\FC837A54.tmp |
| 9 | Data extraction from the source systems Source to analysis area | Data Transfer Icon - Structured |
| 10 | Data is transformed by performing various steps:   1. Data filtering | Vector black filter data icon set. ... | Stock vector | Colourbox |
| 10 | 1. Data cleansing | The Dirty on Data Cleansing & Appending |
| 10 | 1. Data deduplication | C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\7DA8F17.tmp |
| 10 | 1. Data merging | C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\3628E2E0.tmp |
| 10 | 1. Data transposing | Questions from Alteryx Training | InterWorks |
| 10 | 1. Data mapping | C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\1F9097E3.tmp |
| 10 | 1. Data outlier detection | Local outlier factor - Wikipedia |
| 10 | 1. Data imputation, as necessary | Chapter 4 Multiple Imputation | Book_MI.utf8.md |
| 10 | 1. Data aggregation | Download Free png Free Aggregation Icon 203770 | Download ... |
| 11 | Finally getting raw data to structured data.  Some of these steps are repeated on a periodic basis to ensure that the database is maintained at the expected levels of performance. | C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\413B5782.tmp |
| 12 | Some of the data transformations are performed specific for each analysis so that the data is fit for purpose of that specific analysis |  |
| 12 | 1. Hospital management | C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\62C2BC8D.tmp |
| 12 | 1. Individual researcher | Analytics - Free people icons |
| 12 | 1. Health authorities | C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\553C02C8.tmp |
| 12 | 1. Data mining | C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\69771C16.tmp |
| 12 | 1. Various publications | Modern Outline Style Data Analytics Icons Collection Stock ... |

The following visualization provides us with the Flow diagram from data source to final usage by various usage types

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data Sources | Staging area | Ware house | Data marts | Usage |
| C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\9805807.tmp  Operational system  C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\9805807.tmp  Coding dictionaries  C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\9805807.tmp  Clinical system  C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\50D212FE.tmp  Flat files information | C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\9805807.tmp | Data Warehouse Icons - Download Free Vector Icons | Noun Project | C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\9805807.tmp  Operational data  C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\9805807.tmp  Pharmacy data  C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\9805807.tmp  Patient level data | C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\62C2BC8D.tmp  Hospital management  Analytics - Free people icons  Individual vaidya researcher  C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\553C02C8.tmp  Health authorities  C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\69771C16.tmp  Data mining  Modern Outline Style Data Analytics Icons Collection Stock ...  Various types of documents |

Which picture to keep:

Flow diagram from data source to final usage by various usage types

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data Sources | Staging area | Ware house | Data marts | Usage |
| Data access:  Monitor Screen Check Mark Symbol Padlock Data Access Icon — Stock ... | Data Transfer Icon - Structured Vector black filter data icon set. ... | Stock vector | Colourbox The Dirty on Data Cleansing & Appending | C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\1F9097E3.tmp | Chapter 4 Multiple Imputation | Book_MI.utf8.md |  |
| C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\9805807.tmp  Operational system  C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\9805807.tmp  Coding dictionaries  C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\9805807.tmp  Clinical system  C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\50D212FE.tmp  Flat files information | C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\9805807.tmp  Calculations and transformations | Data Warehouse Icons - Download Free Vector Icons | Noun Project  Curated and consistent data storage | C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\9805807.tmp  Operational data  C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\9805807.tmp  Pharmacy data  C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\9805807.tmp  Patient level data | C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\62C2BC8D.tmp  Hospital management  Analytics - Free people icons  Researchers  C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\553C02C8.tmp  Health authorities  C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\69771C16.tmp  Data mining  Modern Outline Style Data Analytics Icons Collection Stock ...  Various documents |
|  | C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\7DA8F17.tmp C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\3628E2E0.tmp  Questions from Alteryx Training | InterWorks | Local outlier factor - Wikipedia | C:\Users\mahajvi1\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\413B5782.tmp |  |

## Challenges

Even in the 15th century, many expressed concerns about the problem of too much information. The 21st centrury scenario of flow of information has been increasing at almost exponential levels, and now it threatens to drown us in data. It poses a major challenge when evaluating its reliability, dependability, while obtaining useful information. Perhaps more significantly, this hurdle probably discourages a large number of people from working with data in the first place. The speed and efficiency of your data prep process directly affects the time it takes to discover insights. American healthcare is in danger of being overwhelmed by data. We undoubtedly require a plan to keep up with this information onslaught. Grasping new knowledge and collecting contemporary information requires more time, work force and money than most providers cannot afford. Right now, the data explosion threatens to destabilize American or rather world medicine.