# 03 Studying demographics and patient specific factors

**Background and Introduction:**

The earlier chapters have explained to us how to collect the data at an observation level and combine these observations into a clinical story. This storyline needs to be read and then understood by different stakeholders in order to fully utilize the data at our disposal. Different scientific areas like demographic analysis, epidemiologic field, real world data, public health perspective, and evidence based medicine have intersections which could be thoroughly understood by the proposed analysis ideas in this chapter.

Integrative TDU hospital is situated in the urban city of Bengaluru, and could be considered a specialty medical center. The modern diseases reported here could picturize the disease burden found elsewhere in the country, but at the same time could have some rare diseases, more number of non-communicable diseases reported, “health-conditions” of not routine nature (need examples from Girish, and other doctors).

There is very little data on the profile of patients accessing traditional systems of medicine. A comparison of profile of patients attending an Ayurveda clinic with that of modern medicine clinic will help in better understanding of utilization of services and preference for system of medicine by the patients seeking health care.

Integration of medicine is a multi-factorial and a complex problem. Is the integrative medical system helping patients or not? Is the system working ok or not? Are the limited medical resources getting efficiently used or not? Who is coming for the treatment and for what kind of medical conditions? What are their health needs? Etc, etc., etc. ... How to use technology, data, methods, and tools to better understand the complex problem at hand? This chapter outlines the proposed methods be used to help a clinician, hospital management, public health policies, epidemiologic views.

The examples defined later in the chapter would generate frameworks for:

1. evidence generation,
2. critical evaluation,
3. efficient storage and recovery,
4. evidence based medicine and
5. evidence synthesis

[https://www.intechopen.com/books/current-topics-in-public-health/clinical-epidemiology-and-its-relevance-for-public-health-in-developing-countries]

**Methods and materials**

Let us identify the basics of various scientific areas like demographic analysis, epidemiologic field, real world data, public health perspective, and evidence based medicine. Tremendous amount of overlap between these scientific streams could be utilized to derive benefits.

Demography:

As per the Australian population association: Demographers study the composition, distribution, and trends of populations. They also make observations about the causes and effects of population changes, such as increases in birth rates or immigration. Demographers collect statistical data, analyze the data to identify any trends, and then predict future trends. These predictions can help governments, social service agencies, and private companies to plan ahead. Demographers are sometimes called population sociologists. Sociology is a broader field than demography and is concerned with the characteristics of social groups.

Roles and functions for demography studies can be broadly defined as:

1. Population projections
2. Inputs into government budget
3. Evidence based policy
4. Communication of vital statistics

Epidemiology:

As per NIH: Epidemiology is the branch of medical science that investigates all the factors that determine the presence or absence of diseases and disorders. Epidemiological research helps us to understand how many people have a disease or disorder, if those numbers are changing, and how the disorder affects our society and our economy. The epidemiology of human communication is a rewarding and challenging field.

Roles and functions for epidemiological studies can be broadly defined as:

1. Incidence: The number of new cases of a disease or disorder in a population over a period of time.
2. Prevalence: The number of existing cases of a disease in a population at a given time.
3. Cost of illness: Many reports use expenditures on medical care (i.e., actual money spent) as the cost of illness. Ideally, the cost of illness would also take into account factors that are more difficult to measure, such as work-related costs, educational costs, the cost of support services required by the medical condition, and the amount individuals would pay to avoid health risks. (Adapted from the Environmental Protection Agency’s Cost of Illness Handbook)
4. Burden of disease: The total significance of disease for society, beyond the immediate cost of treatment. It is measured in years of life lost to ill health, or the difference between total life expectancy and disability-adjusted life expectancy (DALY). (Adapted from the World Health Organization)
5. DALY (Disability-Adjusted Life Year): A summary measure of the health of a population. One DALY represents one lost year of healthy life and is used to estimate the gap between the current health of a population and an ideal situation in which everyone in that population would live into old age in full health. (Adapted from the World Health Organization)

Pharmaco-epidemiology:

As per John Hopkins university: Pharmaco-epidemiology is the study of the utilization and effects of drugs in large numbers of people; it provides an estimate of the probability of beneficial effects of a drug in a population and the probability of adverse effects. It can be called a bridge science spanning both clinical pharmacology and epidemiology.

Pharmaco-epidemiology concentrates on clinical patient outcomes from therapeutics by using methods of clinical epidemiology and applying them to understanding the determinants of beneficial and adverse drug effects, effects of genetic variation on drug effect, duration-response relationships, clinical effects of drug-drug interactions, and the effects of medication non-adherence.

Roles and functions for pharmaco-epidemiological studies can be broadly defined as:

1. Continuous monitoring of patients for unwanted effects and safety concerns
2. Duration response relation
3. Clinical effects of drug-drug interactions
4. Effects of medication non-adherence

Real world evidence / data:

As per US FDA: RWE as “data regarding the usage, or the potential benefits or risks, of a drug derived from sources other than traditional clinical trials”. Real-World Data (RWD) are data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources. Real-World Evidence (RWE) is the clinical evidence about the usage and potential benefits or risks of a medical product derived from analysis of RWD.

Examples of RWD include but not limited to, the following: (1) electronic health records (EHRs); (2) medical claims and billing data; (3) data from product and disease registries; (4) patient-generated data, including from in-home-use settings; (5) data gathered from mobile devices; (6) registries.

Roles and functions for real world evidence studies can be broadly defined as:

1. Generating hypotheses for testing in randomized controlled trials
2. Identifying drug development tools
3. Assessing trial feasibility by examining the impact of planned inclusion/exclusion criteria in the relevant population, both within a geographical area or at a particular trial site
4. Informing prior probability distributions in Bayesian statistical models
5. Identifying prognostic indicators or patient baseline characteristics for enrichment or stratification
6. Assembling geographically distributed research cohorts (e.g., in drug development for rare diseases or targeted therapeutics)

Public health:

As per WHO: Public Health is defined as “the art and science of preventing disease, prolonging life and promoting health through the organized efforts of society” (Acheson, 1988; WHO). Activities to strengthen public health capacities and service aim to provide conditions under which people can maintain to be healthy, improve their health and wellbeing, or prevent the deterioration of their health. Public health focuses on the entire spectrum of health and wellbeing, not only the eradication of particular diseases. Many activities are targeted at populations such as health campaigns. Public health services also include the provision of personal services to individual persons, such as vaccinations, behavioral counselling, or health advice.

Roles and functions for public health studies can be broadly defined as:

1. Implement health educational programs
2. Recommend policies
3. Administer services
4. Conduct research

Pharmaco-epi-real world demographics perspective

Public health

As per WHO: defines “the art and science of preventing disease, prolonging life and promoting health through the organized efforts of society” (Acheson, 1988).

1. Predictions from **demography studies** can help governments, social service agencies, and private companies to plan ahead
2. Factors causing diseases and disorders studies in **epidemiology** are managed by Public health intervention as “Preventive medicine”
3. Insights from **RWD** can be used for disease surveillance and population health
4. **Pharmaco-epidemiology** provides utilization of drugs in large population

Roles and functions:

1. Implement health educational programs
2. Recommend policies
3. Administer services
4. Conduct research

Real world evidence (RWE)

RWE is the clinical evidence, derived from sources other than traditional clinical trials, about the usage and potential benefits or risks of a medical product derived from analysis of RWD, from a variety of sources

Roles and functions:

1. Generate new hypotheses
2. Identify drug development tools
3. Generate probability functions for statistical models
4. Identify prognostic factors / baseline characteristics for cohort studies
5. Assess new trial feasibility based on RWD evidence

Pharmaco-epidemiology

A bridge between clinical pharmacology and epidemiology to study the utilization and effects of drugs in large numbers of people

Role and functions, to estimate:

1. The probability of beneficial effects of a drug
2. The probability of adverse effects
3. Effects of genetic variation on drug effect
4. Duration-response relationships
5. Clinical effects of drug-drug interactions
6. Effects of medication non-adherence

Epidemiology (focused on micro level – each patient)

1. Investigates all the factors that determine the presence or absence of diseases and disorders
2. Helps to understand how many people have a disease or disorder, if those numbers are changing, and how the disorder affects society and economy

Roles and functions, to calculate:

1. Incidence rate
2. Prevalence rate
3. Cost of illness
4. Burden of disease
5. DALY (Disability-Adjusted Life Year)

Demographics (focused on macro level - population sociology)

1. Study the composition, distribution, and trends of populations
2. Collect statistical data, analyze the data to identify any trends, and then predict future trends
3. Observations about the causes and effects of population changes

Roles and functions:

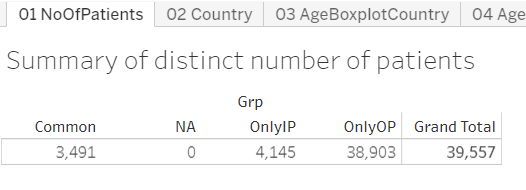
1. Population projections
2. Inputs into government budget
3. Evidence based policy
4. Communication of vital statistics

Results:

This section outlines example analyses carried out to describe the clinical story emphasizing the importance of each of the streams defined above:

Actual presentations to be included in this chapter:

1. Total number of patients



Data version: 2011 to Oct 2016

<https://public.tableau.com/views/04_patient_analysis_tablaeu/05bNoOfDis_agebox?:display_count=y&:origin=viz_share_link>

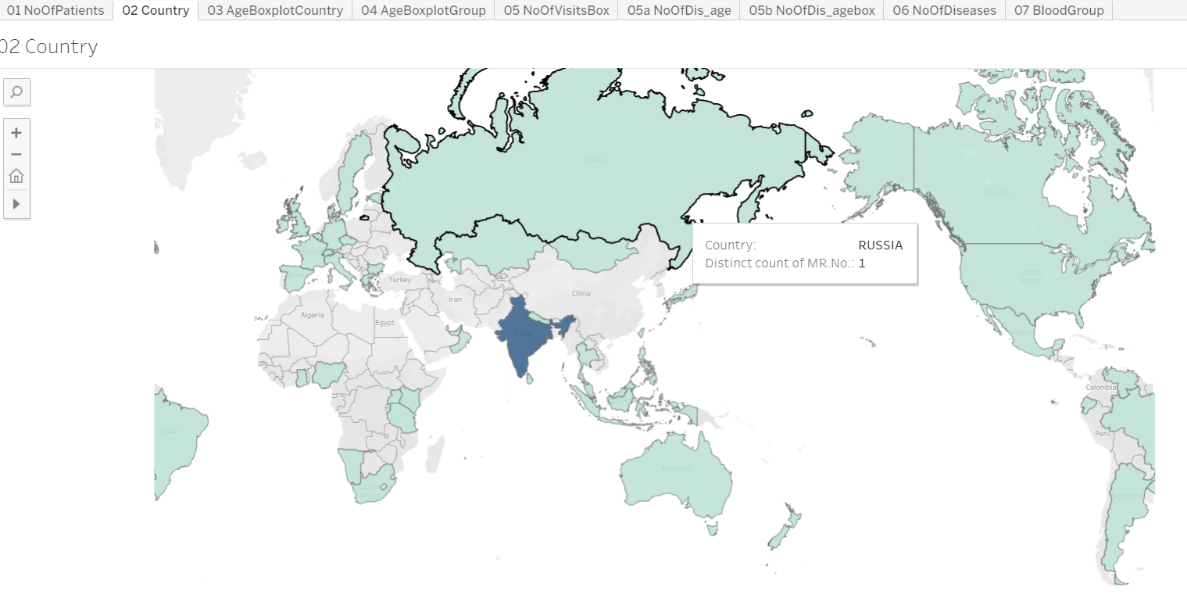
Interpretation:

1. The very first analysis to get a sense of size of the patient population available.
2. This version of the data has ~40,000 patients, majority of the patients, 90%, are Out Patients and 10% of the patients are In-patients.
3. This analysis shows that the daily visits of the patients must be suitably managed and would need appropriate staffing to cater to patients.
4. We do not have data for “waiting time to see the doctor” – but if this is made available then we can get very good insights into one part of every-day experience a patient goes through before meeting a doctor.

Use for each stakeholders:

This example can be considered as a work from demography stream, providing the necessary tabulations. These numbers provide empirical evidence about the quantum of patients. If extrapolated at the sub-district, district level, they can form the basis of public health policies framed either by government or by private companies.

1. Country wise visualization



Data version: 2011 to Oct 2016

<https://public.tableau.com/views/04_patient_analysis_tablaeu/05bNoOfDis_agebox?:display_count=y&:origin=viz_share_link>

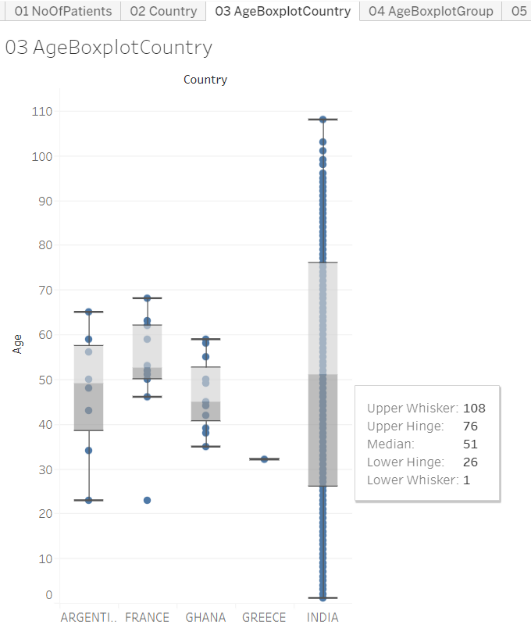
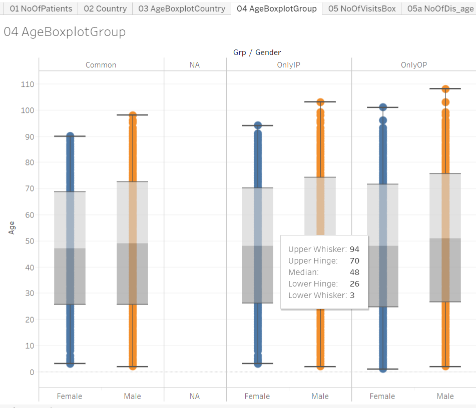
Interpretation:

1. This graphic shows the geographical distribution of patients.
2. In this version of the data, there are patients coming from 50+ different countries. 90% or more patients are from India and remaining 10% patients are from different parts of the world.
3. If the individual state and city information is available then additional drill down representation is possible – this supplementary pictorial will allow us to identify the distribution of patients and diseases from different parts of India.

Use for each stakeholders:

A pictorial representation of the world data, very useful as an executive summary. This form of data representation will help any public health official. More details related to diseases, treatments, additional demographic characteristics could be added to the tooltip to efficiently recover key information as and when needed.

1. Age and gender distribution

Data version: 2011 to Oct 2016

<https://public.tableau.com/views/04_patient_analysis_tablaeu/05bNoOfDis_agebox?:display_count=y&:origin=viz_share_link>

<https://public.tableau.com/views/04_patient_analysis_tablaeu/05bNoOfDis_agebox?:display_count=y&:origin=viz_share_link>

After understanding the data about the number of patients and where do they come from, it will be logical to discover the underlying distribution by different categorical factors like age, gender, disease types, prakriti types, etc. Are older people coming to the hospital, are more females coming compared to males? How long will they continue coming, are answered through these data analyses.

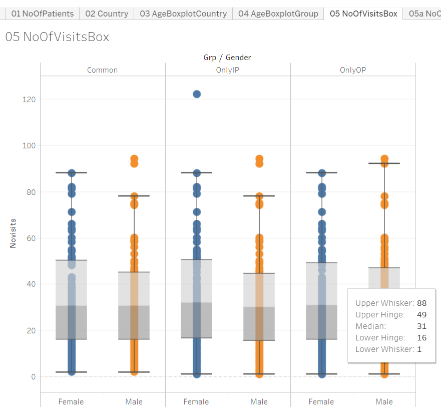
Interpretation:

1. The above 2 pictorials provide age distribution by different countries and by gender.
2. Median age for females is marginally higher than males across all visit types.

Use for each stakeholders:

Example of RWE to provide a quick snapshot of health seeking behaviors. This evidence generated using real world data could be further synthesized to understand the underlying clinical scenario.

1. Number of visits, and visit types:



Data version: 2011 to Oct 2016

<https://public.tableau.com/views/04_patient_analysis_tablaeu/05bNoOfDis_agebox?:display_count=y&:origin=viz_share_link>

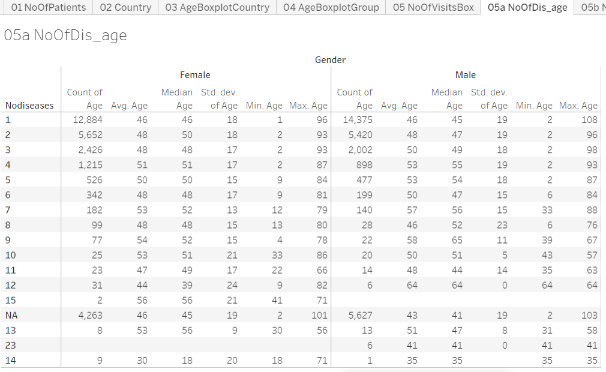
Interpretation:

1. The above 2 pictorials provide age distribution by different types of visits and by gender.
2. Median age for females is marginally higher than males across all visit types.
3. There are a few outliers observed for the In Patient visits for both the genders, with females having more number of visits.

Use for the stakeholders:

The first example showed us that there are fewer number of In-patients, compared to the out patients. This provides great empirical insights on the potential group of patients visiting the hospital. The diseases however severe as well as chronic do not appear to be life threatening with imminent deaths. Possible evidence synthesis could be that the hospital management seem to be Ok with patients not opting for In-patient services.

1. Number of diseases by age and gender



Data version: 2011 to Oct 2016

<https://public.tableau.com/views/04_patient_analysis_tablaeu/05bNoOfDis_agebox?:display_count=y&:origin=viz_share_link>

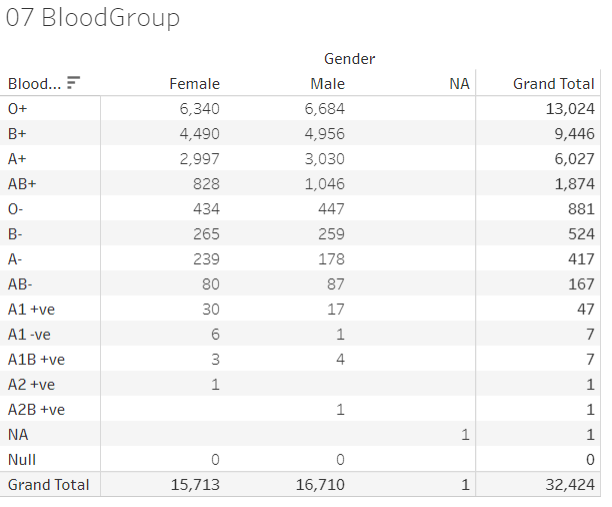
Interpretation:

1. The above table summarizes number of diseases reported by gender.
2. Almost 12k+ and 14k+ patients have reported only a single disease, these could have come only once to the hospital and may not have come back at all after reporting the 1st disease. The median age of 46 years is same for both the genders. Are these patients largely coming in for “2nd opinion”? Or if this data is to be looked at positively, they are getting benefitted and hence are not coming back for consultation beyond the first reported disease?
3. The summary statistics for each of the category across number of diseases looks quite similar.
4. There are a few outliers observed having more than 10 disease conditions across the years. These patients could be having a lot of faith in Ayurvedic treatment. For these patients to continue on, they could have found the underlying treatment effective.
5. The maximum age of the 108 is a possible case of data issue. While finding data issue was not a primary outcome of the analysis, there is this secondary usage available to the scientific community.

Use for the stakeholders:

This empirical evidence will be very useful for the hospital management, public health officials, treating physicians. This kind of tabulation plays a key role in evidence generation and synthesis. Is there a similar tabulation available for another Ayurvedic hospital, or any other private or public hospital? Is there a similar distribution viewed? This evidence can be used to understand the use and misuse of the limited medical sources across the geographies.

1. Blood group distribution

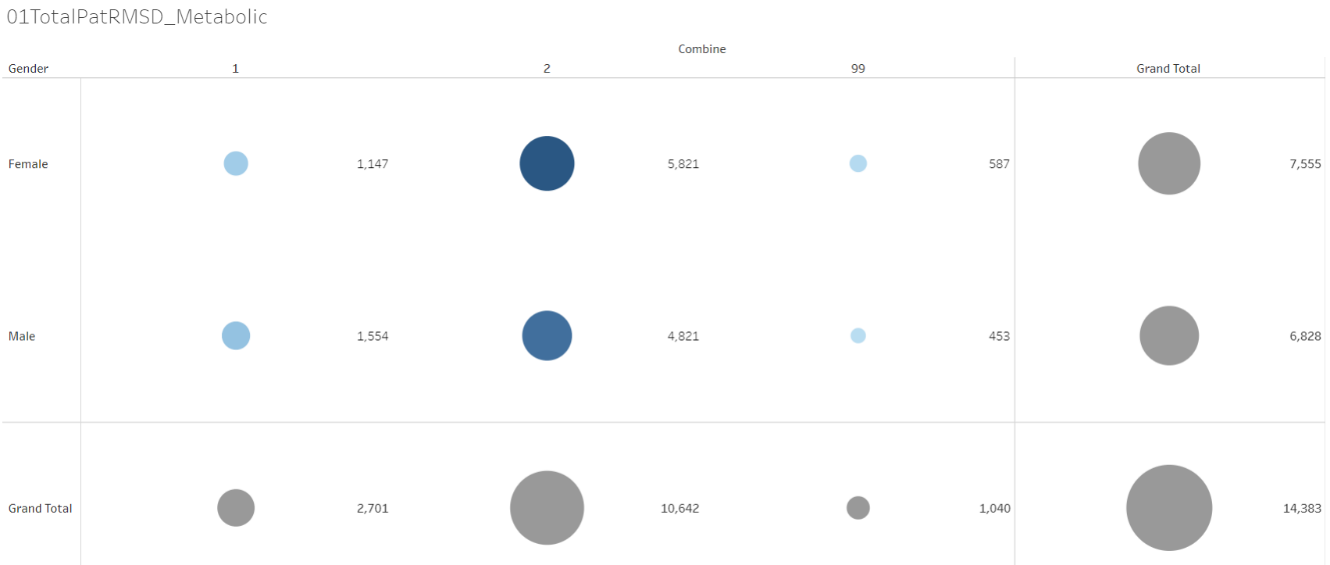


Data version: 2011 to Oct 2016

<https://public.tableau.com/views/04_patient_analysis_tablaeu/05bNoOfDis_agebox?:display_count=y&:origin=viz_share_link>

Interpretation and use for stakeholder:

1. Blood group distribution for such a large number of patients is a great source of knowledge. Even though this does not help in every day treatment options, there is undoubted epidemiological value in this tabulation.
2. There are obvious mistakes in documenting the blood groups observed via this tabulation – another secondary use of this tabulation to build data quality related efficiencies.
3. These analyses carried out for all patients, RMSD, metabolic, CKD, Cancer groups



Data version: 2011 to Oct 2016

<https://public.tableau.com/views/01RMSD_MET/01TotalPatRMSD_Metabolic?:display_count=y&:origin=viz_share_link>

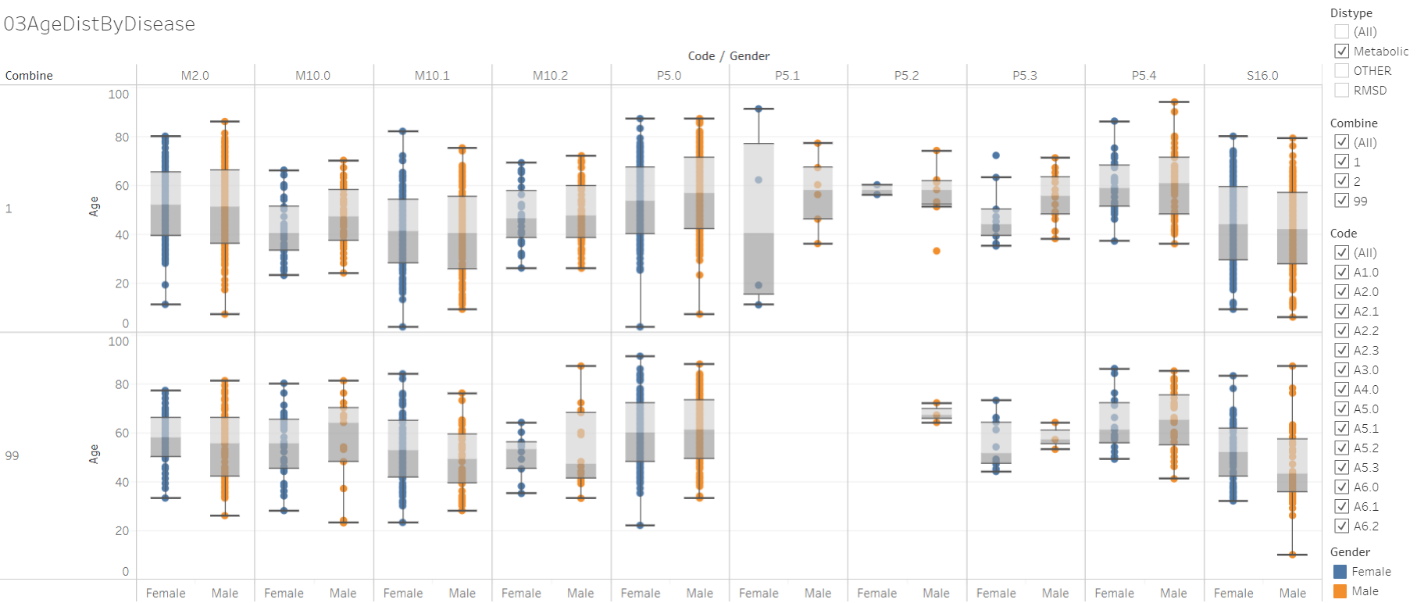
Interpretation:

1. The above bubble plot shows the number of patients in the 2 disease groups.
2. It is quite clear that there are a lot more patients in the RMSD group compared to the metabolic group.
3. Traditionally ayurveda has been used to treat chronic vata vyadhi, which is documented here by the underlying data.

Use for the stakeholders:

Useful macro level representation of data for public health policies.

1. Different diseases as per age groups



Data version: 2011 to Oct 2016

<https://public.tableau.com/views/01RMSD_MET/01TotalPatRMSD_Metabolic?:display_count=y&:origin=viz_share_link>

Interpretation:

1. The above visualization provides age group distribution by gender and disease.
2. The metabolic diseases are displayed above.
3. This picture provides a bird’s eye view on the age distribution via plotted boxplots.
4. The above picture summarizes data for more than 3000 patients.
5. Gender wise differences in treatment duration



Data version: 2011 to Oct 2016

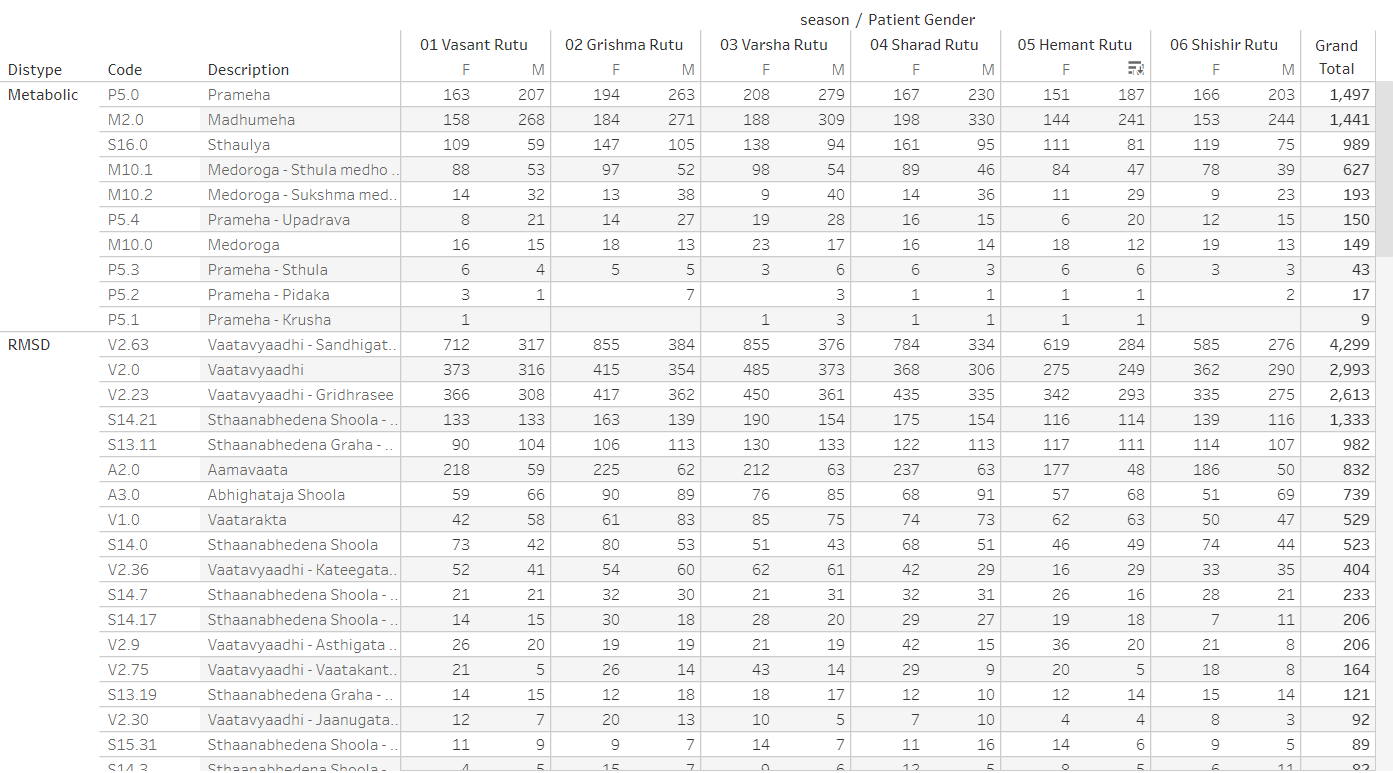
<https://public.tableau.com/views/01RMSD_MET/01TotalPatRMSD_Metabolic?:display_count=y&:origin=viz_share_link>

The duration between the very first visit and the last visit has been calculated. The duration then categorized into the following way:

(1) >= 1 day, (2) >= 1 month, (3) >= 2 months, (4) >= 3 months, (5) >= 6 months, (6) >= 1 year, (7) >=2 years, (8) >= 3 years, (9) >= 4 years and (10) >= 5 years.

Interpretation:

1. Out of the 14383 patients, approximately 38% patients come to hospital for more than 1 month. Only 15% patients visit the hospital after 1 year.
2. If the patients are getting benefit of the treatment and due to that patients are not coming back then that is a great advertisement, but if the treatment benefit is not experienced and hence they do not come then that is not a situation to be in.
3. Patients only having RMSD or metabolic diseases seem have to similar percentage of “retention”, by the patients with RMSD as well as metabolic diseases seem to come back more. (>=1 year: ~15% vs. 32%)
4. Seasonal variations along with other factors



Data version: 2011 to Oct 2017

<https://public.tableau.com/views/01SQL_Dis_Med_Ser/MedicineByDay?:display_count=y&:origin=viz_share_link>

Rate of occurrence for diseases for metabolic and RMSD: after understanding the data and some operational side of it, let us understand the rate of occurrence for various diseases reported. The most frequently reported diseases have been sorted in ascending order.

1. Metabolic disease group has 10 diseases and RMSD disease group has 100+ diseases.
2. For Metabolic disease group the top 3 diseases are:
   1. Prameha
   2. Madhumeha
   3. Sthaulya
3. For RMSD disease group the top 5 diseases are:
   1. Vaatavyaadhi – Sandhigata Vaata
   2. Vaatavyaadhi
   3. Vaatavyaadhi – Gridhrasee
   4. Sthaanabhedana Shoola – Katee Shoola
   5. Sthaanabhedana Graha – Katee Graha

Interpretation and use for the stakeholders:

1. Prameha and Madhumeha are reported more by males than females.
2. There are more female patients with disease condition Sthaulya.
3. In general, RMSD diseases are reported by more number of females than males.
4. For RMSD disease group, 51 out of 100+ disease are reported by <= 10 patients.
5. Pre and post disease classification analysis

This data is generated from every day medical practice at the hospital. Hence the diseases are reported almost at random. The following analysis uses 1st occurrence of any disease as day 1 at an individual patient basis. Using this as a reference day “before period” and “after period” is derived. “Before period” provides significant amount of “baseline data”, “after period” provides specific insights into what would happen after the onset of the reference disease.

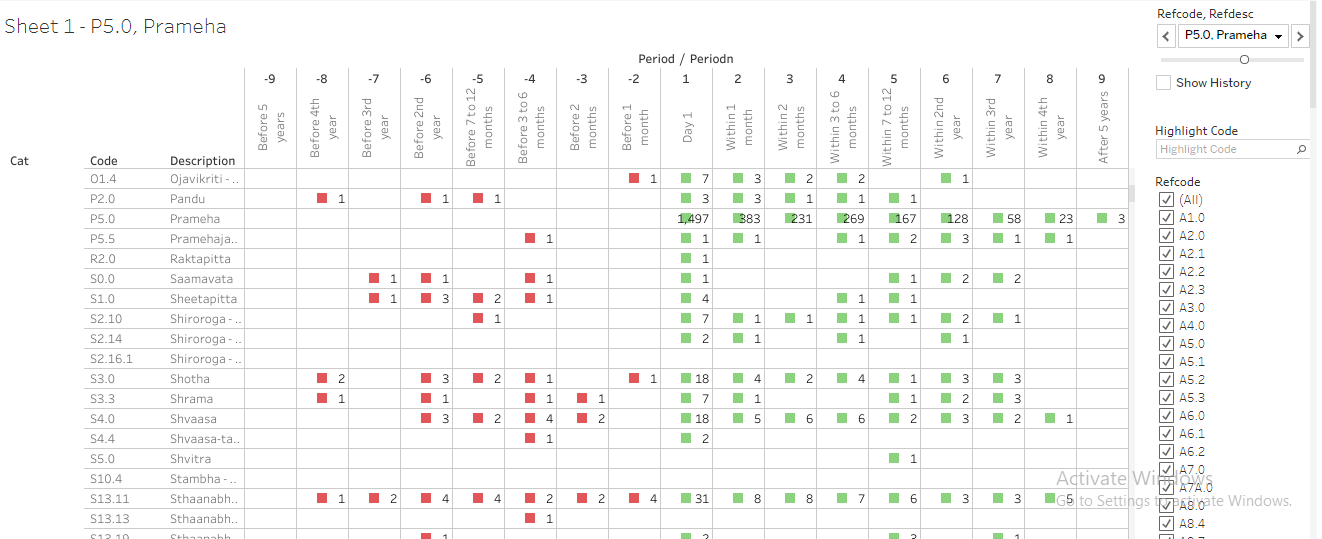
Algorithm to create the underlying data for analysis:

1. Each of the 106 diseases (10 Metabolic and 96 RMSD) is considered as a reference disease.
2. Day 1 is calculated as the reference day 1 for individual patient for each disease.
3. Other diseases for the same patient are positioned either before or after compared to this reference disease.
4. Duration w.r.to this reference day is calculated before and after day 1. This calculation provides the background view as well as future view.
5. This referencing allows for more informative background disease as well as background medicine information. The duration is split into the following time points:

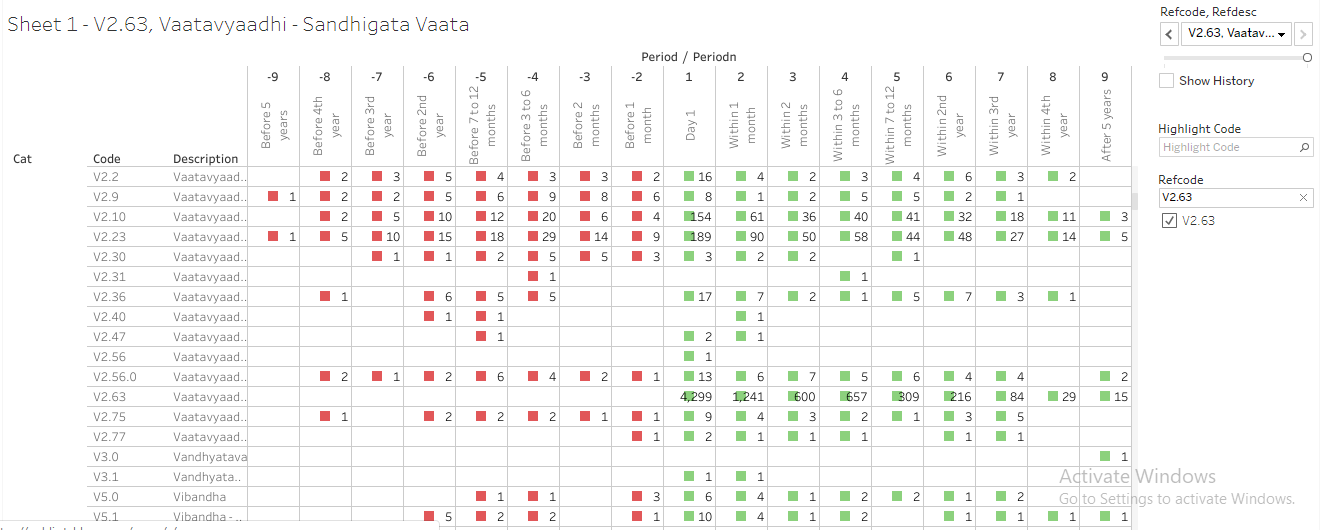
| **Before** | **After** |
| --- | --- |
| Day 1 as reference |  |
| Before 1 month | Within 1 month |
| Before 2 months | Within 2 months |
| Before 3 to 6 months | Within 3 to 6 months |
| Before 7 to 12 months | Within 7 to 12 months |
| Before 2nd year | Within 2nd year |
| Before 3rd year | Within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

E.g. use Prameha as an example:

1. Prameha has been reported by 1497 patients. Out of these 383 patients visit hospital within 1st month, 231, 269, 167, 128, etc are in the following time points.
2. Other lines in the table provide details about diseases reported by these 1497 patients.
3. Bottom section of the table provides information about the treatment details for these patients.



Another example of Vaatavyadhi – Gridhrasee, 4299 patients have reported at least once.



Data version: 2011 to Oct 2017

<https://public.tableau.com/views/085_dis_1st_time_refCal_NodesEdges/Sheet1?:display_count=y&:origin=viz_share_link>

1. Prameha has been reported by 1497 patients. Out of these 383 patients visit hospital within 1st month, 231, 269, 167, 128, etc are in the following time points.
2. Other lines in the table provide details about diseases reported by these 1497 patients.
3. Bottom section of the table provides information about the treatment details for these patients.

Details about the analysis file:

1. 1 sheet for each reference disease is created.
2. Frequency count of diseases and prescribed medicines is displayed.
3. “Before period” counts are displayed in red colour and “After period” counts are displayed in Green colour. This analysis provides good insights into the causal relationships.

Interpretation and use for the stakeholders:

1. Many patients drop off after the very first day. This could be seen in 2 ways positive view: the patients are getting immediate relief and hence not coming back. Negative view: patients only came in for 2nd opinion and did not want to come back after the 1st interaction.
2. The before and after visualization of data allows to build a disease trajectory to be used by a doctor
3. Possible diagnostic and prognostic relationships could be created and visualized contributing to public health research.
4. Demographics analysis using the derived dataset

This section provides insights into patient level data created above. Full analysis is stored in another HTML file. The analysis is split into 7 different sections as follows.

Data version: 2011 to Oct 2017

The link to the report: <https://rpubs.com/mahajvi/390061>

1. Frequency counts to understand patient population
   1. Disease category

Patients are classified into 2 disease groupings, namely metabolic and RMSD diseases. There are 10 diseases contributing to metabolic and 106 diseases contributing to RMSD group.

* 1. Disease category by gender

There are 1343 females and 1771 males having metabolic diseases. There are 7180 females and 5778 males having RMSD diseases. There are more numbers females than males for RMSD diseases.

* 1. Number of metabolic patients

There are 4447 patients with metabolic disease.

* 1. Number of RMSD patients

There are 14292 patients with RMSD diseases. There are 1333 patients with both metabolic and RMSD diseases.

1. Summary statistics of age at baseline and subsequent visits
   1. Metabolic: Summary statistics of baseline age in years

For females Mean (SD) is observed as 46.5 (14.51) years, for males, it is 49.4 (13.96) years.

* 1. RMSD: Summary statistics of baseline age in years

For females Mean (SD) is observed as 48.8 (14.70) years, for males, it is 47.7 (15.88) years.

* 1. Metabolic: Summary statistics of age in years, by visit

The mean and median age for female patients go on increasing from 46 to 57, similarly for males it is 49 to 52.

* 1. RMSD: Summary statistics of age in years, by visit

The mean and median age for female patients go on increasing from 48 to 57, similarly for males it is 47.7 to 52.

1. Analysis related to number of visits and duration
   1. Metabolic: Summary statistics in days and visits

* Total duration of visits to hospital for females Mean (SD) is 265.2 (448.15) days, median (range) is 34 (1-2506) days, for males Mean (SD) is 272.1 (472.95), median (range) is 30 (1-2506) days.
* Total number of IP visits to hospital for females Mean (SD) is 6.3 (5.82) visits, median (range) is 5 (1-39) visits, for males Mean (SD) is 6.9 (5.04), median (range) is 5 (1-60) visits.
* Total number of OP visits to hospital for females Mean (SD) is 5.3 (8.86) visits, median (range) is 2 (1-230) visits, for males Mean (SD) is 5.4 (12.13), median (range) is 2 (1-318) visits.
  1. RMSD: Summary statistics in days and visits
* Total duration of visits to hospital for females Mean (SD) is 234.0 (431.44) days, median (range) is 21 (1-2528) days, for males Mean (SD) is 222.3 (441.60), median (range) is 10 (1-2530) days.
* Total number of IP visits to hospital for females Mean (SD) is 7.0 (5.49) visits, median (range) is 6 (1-86) visits, for males Mean (SD) is 7.5 (7.02), median (range) is 6 (1-83) visits.
* Total number of OP visits to hospital for females Mean (SD) is 4.8 (8.96) visits, median (range) is 2 (1-274) visits, for males Mean (SD) is 4.3 (9.47), median (range) is 2 (1-318) visits.
  1. Diseases: Summary statistics in days, by gender

This section provides summary statistics by gender for each disease.

1. Cumulative analysis: In this analysis patients are counted multiple times as per available data for each time period. Following time points are considered for analysis: Day 1, >=1 month, >=2 months, >=3 months, >=6 months, >=1 year, >=2 years, >=3 years, >=4 years and >=5 years. These provide clinical and operational insights into disease manifestations. A patient visiting for more than 5 years is counted in all categories. If a patient has discontinued in the 4th month then that patient is counted in Day 1, >=1 month, >=2 months, >=3 months categories.
   1. Cumulative display of patients by duration

* Total 17406 patients visit hospital on day 1. Out of these 8725 (50%) continue hospital visits after 1 month, 7018 (40%) continue visits after 2 months, 6173 (35%) continue visits after 3 months, 4785 (27%) continue visits after 6 months, 3446 (19%) continue visits after 1 year. 2020 (11%) have visits beyond 2 years, 1168 (6%) have visits beyond 3 years, 579 (3%) have visits beyond 4 years and 256 (1.5%) have visits beyond 5 years.
  1. Cumulative display of patients by duration and gender
* Similar patterns for both RMSD and metabolic diseases by gender are observed. Some diseases are cured only after 1 visit. For some diseases, approximately 5 visits are sufficient, for some diseases more than 5 visits spanning more than 30 days, 60 days, 90 days, etc. are needed. Additional analysis is carried out to explore the data.
  1. Cumulative display of patients by Code and duration
* Analysis for each disease is carried out. The drop-out pattern for each of the diseases is consistent with overall duration analysis.

1. Non-overlapping analysis: this analysis provides information for different time points in mutually exclusive manner. An individual patient is counted only once for each duration period.
   1. Total patients present across different time points
   2. Summary statistics of total duration across different time points
   3. Summary statistics of total visits across different time points
   4. Summary statistics of total visits across different time points for each disease
   5. Total duration across non overlapping time periods
   6. Metabolic: total duration by gender across non overlapping time periods
   7. RMSD: total duration across non overlapping time periods
   8. Metabolic: total duration for by gender across non overlapping time periods
   9. Metabolic: total duration for each disease across non overlapping time periods
   10. RMSD: total duration for overall by gender non overlapping time periods
   11. RMSD: total duration for each disease across non overlapping time periods
2. No-overlapping time period, frequency counts
   1. Frequency counts for Total number of patients with treatment and diseases
3. Diseases present in each non-overlapping duration: If a disease is present at least in a time period once then denote it by Yes.

* This table does not provide the quantum of patients reporting diseases across time points, but only provides binary representation of presence or absence of a disease.
* This analysis shows, frequency of a disease getting reported across different time periods. If a disease is presented in all the time point categories then it means that the disease is reported consistently. Speculative interpretation could be “patients are getting some benefit”.
* Some diseases which are reported first time only in say after 4th month, or 1st year, etc. could be “additional co-morbities” developed in due course of primary disease. Other interpretation could be that these could be side effects of prescribed treatments.

References to the analysis files:

|  |  |
| --- | --- |
| R program | 100\_adsl\_analysis.RMD   * 07\_cumulative\_duration.R * 07\_cumulative\_dur\_byCode.R * 07\_cumulative\_dur\_byCode\_Part02.R * 08\_nonoverlap.R |
| Datafile | 01adsl\_met\_rmsd.rds |
| KnitR output | 100\_adsl\_analysis.HTML |
| Tableau vizname | 01SQL\_Dis\_Med\_Ser |
| Tableau sheetname | 1. RMSD\_Met\_patients:    * Frequency table by gender and high level disease classification, there are more number of RMSD patients compared to the Metabolic, Metabolic and RMSD patients.    * There is more number of female RMSD patients compared to males.    * There are similar number of males and females in Metabolic disease categories 2. Visit\_Duration:    * Boxplot is plotted for Total duration of hospital visits is calculated as the maximum date of hospital visit - minimum date of hospital visit + 1 in days for each patient, by gender and disease group |

Interpretation:

1. There is more number of RMSD disease patients compared to metabolic disease patients.
2. For metabolic disease group Males and females at baseline have similar age characteristics.
3. For RMSD disease group median age of females is more than median age of male patients.
4. Approximately 50% of patients come only for 1 visit.
5. As number of visits increase, the median age for both males and females continue increasing, elder patients are seen continuing for more number of visits.
6. Maximum duration to hospital is almost 2500+ days; median number of days is approximately 30+ days.
7. Maximum number of IP visits is 39 and 60 (female and male), maximum number of OP visits is 230 and 318 (female and male).

Discussion:

The above examples provide good insights into varied components feeding into public health domain. It is more and more acknowledged that evidence produced using “real-world data” (RWD) is vital for gauging the safety and effectiveness of health-related interventions. Basic science can use the relations developed through the above analysis for disease and generation of disease categorizations. Public health domain can be benefitted by disease surveillance and population health. Clinical research initiatives will get benefited by having access to data generated in heterogeneous sets of people.

**The focus of each of the scientific stream contributes to Public health needs:**

The health and healthcare requirements of a population cannot be calculated without understanding of its magnitude and characteristics. Demography is concerned with this essential ‘numbering of the people’ and with understanding population dynamics—how populations change in response to the interplay between fertility, mortality, and migration. This understanding is a pre-requisite for making the forecasts about future population size and structure, which should underpin healthcare planning. Analysis of both the present and the future necessitates a review of the past. [Oxford]

1. Demography study: focuses on macro level (aggregate level) size and structure
2. Epidemiology study: focuses on diseases affecting the individuals (at micro level)
3. Pharmaco-epidemiology: study focuses on prescription drug use and its consequences in populations, continuous monitoring of patients for unwanted effects and safety concerns
4. Real world evidence study: focuses on non-traditional data for secondary uses

The above mentioned streams provide inputs into Public health which relies upon scientific techniques needed to reduce the incidence of diseases and their origins. Public health focuses on the entire spectrum of health and wellbeing, not only the eradication of particular diseases. Many activities are targeted at populations such as health campaigns. Public health services also include the provision of personal services to individual persons, such as vaccinations, behavioral counselling, or health advice.

Some of the tools, approaches outlined above can be utilized.

|  |
| --- |
| Summary of the chapter: This chapter   * Outlines the clinical understanding which can be derived from the existing data. How individual observations can be transformed into meaningful stories at hospital * How can concepts from scientific areas like demographic analysis, epidemiologic field, real world data, public health perspective, and evidence based medicine, be combined * Provides rich empirical evidence * Actionable inputs to hospital management, practicing doctors and for research publications * Lays down the foundation for the next chapter of “understanding diagnostic factors” |

These analyses are planned but not yet completed:

(12) Data outlining jobs profile of patients — see if that analysis is still present

(13) Treatment assignment as per age - paediatric, adult, geriatrics

(14) DALY score calculations and related analysis - not done yet

(15) Use this analysis in defining new studies as evidence based background