Use of databases and software tools to build empirical evidence of Ayurvedic practice

Vinay Mahajan\*1, 2, Ashwini Mathur 3, Girish Tillu1

\* and 1: Corresponding author, TDU, 2: Novartis Pharmaceuticals Pvt. Ltd.

Postal Address: Institute of Transdisciplinary Health Sciences and Technology (TDU)

C/o Foundation of Revitalization of Local Health Traditions (FRLHT)

#74/2, Jarakabande Kaval, Post Attur via Yelahanka, Bangalore. PIN: 560 064, India.

2: Novartis Pharmaceuticals Pvt. Ltd. Raidurgam, Serilingampally, Hyderabad PIN: 500081, India.

3: Novartis Ireland, Vista Building, Elm Business Park, Merrion Road, Ballsbridge, Dublin 4

Phone number: +91 99 66 999 523.

Abstract:

Key words: R programming, Tableau, Visual displays, Ayurvedic data, empirical evidence

Introduction and background:

Use of technology since early 1970s allowed healthcare industry to build patient level databases [1, 2]. This provided insights into the medical practice, disease progress, correlations between diseases, causal relationships, etc [3, 4, 5, 6, 7]. This was made possible due to the systematic collection of data, cleaning, creating international standards for coding free textual data, then tabulating and analyzing [8, 9, 10]. These in turn gave rise to development of databases, statistical software and data visualization software [11, 12].

Ayurvedic medical practice is believed to be very old and has a large supporting data in form of classical texts, pharmacopeias, anecdotal evidence as well as flourishing medical colleges, medical practice and scientific journals [13, 14, 15, 16, 17, 18]. One area where it could lack behind is having a large scale patient level database documenting diseases, methods of treatment, outcomes and safety, efficacy profiles of prescribed drugs, Ayurvedic procedures [19, 20, 21].

Ayurvedic hospital in TDU is operational since year 2011 and perhaps hosts one of the largest electronic databases of Ayurvedic medical practice [22]. The hospital database contains clinical data for more than 51000 patients since 2011 to October 2017. The database contains complete patient information. The Ayurvedic disease classification dictionary (ACD) is used to code the diseases [Dictionary developed at the University, and CDAC, Pune]. There are approximately 150,000 visits recorded, covering more than 900 unique diseases, more than 3000 medicinal procedures\*, in patient as well as outpatient visits [Based on the hospital database analysis]. The scale of the data and no artificial limitations of inclusion and exclusion criteria set of clinical should provide real world evidence view or the epidemiological view for the above stated period.

Basic understanding of the demographic, background history, visit pattern, treatments prescribed at such a large scale should help in bridging the perceived gap of missing empirical evidence.

Database setup and softwares:

The Ayurvedic hospital should be considered as a pioneer in building electronic health records. The database captures complete picture of patient experience in day to day clinical practice. It captures modern medical parameters obtained through various blood tests, scans, ECGs, etc. as well as it captures Ayurvedic parameters in carefully designed Case Report Forms. The source data is secured in a controlled environment with SQL database at the back end. This data is accessed using SQL queries and then using R programming language for necessary calculations [23, 24]. Summarization and tabulation are carried out using R libraries, visuals are built using tableau software for a very easy to use end user experience. The following steps were followed to convert an individual observation to a meaningful table or a visual (References in supplementary material, Fig 1, 2, 3, and 4). The source data is captured in multiple data tables or views, and these need to be combined or aggregated into meaningful data tables for further processing. To get to a consolidated Data version of 25 variables, 15 different data tables had to be processed.

R is a programming language and free software environment for statistical computing and graphics that is supported by the R Foundation for Statistical Computing. The R language is widely used among statisticians and data miners for developing statistical software and data analysis [24].

Tableau (French for 'little table' literally, also used to mean 'picture'; pl. tableaux or, rarely, tableaus) is a software providing visual query language interface to convert large amounts of data into beautiful interactive graphical displays [25, 26].

All these softwares are freeware softwares or open source, hence there is no additional financial burden on researchers to procure the best in class analytical capabilities.

Study design and analysis, results:

Various scientific and operational components of the database were analyzed and will be listed one by one. The analysis is performed on various cuts of data, (1) analysis on data from 2011 to 2016 for approximately 40000 patients, (2) analysis performed on metabolic and RMSD patients’ areas on data from 2011 to 2017.

Tableau display: this interactive working sheet provides a bird eye view about the basic data. Total number of patients, which country do they belong to, box plots of age by country as well as by gender, number of visits per patient, blood group, number of diseases reported by individual patients (Fig 5, 6, 7 and 8).

Operational view: how many patients are visiting each day, number of new patients, repeat visits, in patient visits, outpatient visits, etc. The following figure shows that more number of patients visit on weekends (Fig 9). Another view shows frequency of patients reporting a specific disease by gender by each month (Fig 10). This view provides a glimpse into seasonal variations.

Assume that the collected data covers CRFs of different kinds, and if one wants to track which data is captured on which visit then the following visual would be very useful (Fig 11).

Patient profile view: this view shows the longitudinal disease and medical history at a glance on one page (Fig 12).

Disease and treatment view: this visual provides a patient level view of number of diseases reported for the first time and then repeated, similarly treatment prescribed for the first time vs. a repeat of treatment. When a new disease is reported, usually a new treatment or treatments are reported, if there is only a new treatment added then it could indicate, the earlier treatment may not have worked, or it explains the treatment regimen (Fig 13).

Box plot for Differences in days from one visit to other by disease. This visual is operationally useful as well as scientifically (Fig 14). In a hospital setting, there are usual 50% drop off rate after the first visit for a patient. Depending upon a chronic disease, acute disease, easily curable condition or a disease requiring longer follow up, the frequency of visits would be shaped.

Conclusion: the above discussion provides simple steps to generate huge volumes of rich data at a minimal cost. The databases could be built at all India level, at each hospital level or at an individual practice level. Even if 10 patient data collection is done each day then this approach has a potential of generating 3000+ patient data each year as a lower end of estimate. This approach should help in generating data to fill the perceived gap of "no empirical evidence". First ever estimates of key Ayurvedic parameters could be attempted. Insurance companies could be big consumers of this type of data.

There is a need to build standard dictionaries covering Ayurvedic parameters to make the data uniform. ACD is one such example of available dictionary. Wherever possible encourage use of ICD, ISO, LOINC, CDISC dictionaries as they are freely available. Standard Ayurvedic drug dictionary should be put in place in case there is nothing available. If an organization has necessary funds then there are licensed dictionaries available at a price.

Conflict of Interest: No conflict of interest.

Financial support: No.

Acknowledgements: TDU University and IAIM hospital, Bengaluru, India.

Authors: All research done by the authors listed on page number 1.

Acronyms:

TDU: Trans Disciplinary University

ACD: Ayurvedic Classification of Diseases

ICD: International Classification of Diseases

ISO: International Organization for Standardization

LOINC: Logical Observation Identifiers Names and Codes

CDISC: Clinical Data Interchange Standards Consortium

Reference:

[1] Michael Fitzmaurice J, Adams K, Eisenberg JM. Three Decades of Research on Computer Applications in Health Care: Medical Informatics Support at the Agency for Healthcare Research and Quality. Journal of the American Medical Informatics Association : JAMIA. 2002;9(2):144-160. doi:10.1197/jamia.M0867.

[2] Institute of Medicine (US) Committee on Improving the Patient Record; Dick RS, Steen EB, Detmer DE, editors. The Computer-Based Patient Record: Revised Edition: An Essential Technology for Health Care. Washington (DC): National Academies Press (US); 1997. 1, Introduction. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK233055/>

[3] Rubin DB. Estimating Causal Effects from Large Data Sets Using Propensity Scores. Ann Intern Med. ;127:757–763. doi: 10.7326/0003-4819-127-8\_Part\_2-199710151-00064

[4] William J. Frawley, Gregory Piatetsky-Shapiro, and Christopher J. Matheus, AI Magazine, Knowledge Discovery in Databases: An Overview, Vol 13 No 3: Fall 1992, DOI: <https://doi.org/10.1609/aimag.v13i3.1011>

[5] Van den Akker M., Buntix F., Metsemakers J.F.M., Roos S., Knottnerus J.A., Multimorbidity in general practice: Prevalence, incidence, and determinants of co-occurring chronic and recurrent diseases, (1998) Journal of Clinical Epidemiology, 51 (5) , pp. 367-375.

[6] Alexia Giannoula, Alba Gutierrez-Sacristán, Álex Bravo, Ferran Sanz & Laura I. Furlong, Identifying temporal patterns in patient disease trajectories using dynamic time warping: A population-based study, Scientific Reports volume 8, Article number: 4216 (2018)

[7] Jensen, A. B., Moseley, P. L., Oprea, T. I., Ellesøe, S. G., Eriksson, R., Schmock, H., … Brunak, S. (2014). Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. Nature Communications, 5, 4022. <http://doi.org/10.1038/ncomms5022>

[8] Krishnankutty, B., Bellary, S., Kumar, N. B. R., & Moodahadu, L. S. (2012). Data management in clinical research: An overview. Indian Journal of Pharmacology, 44(2), 168–172. <http://doi.org/10.4103/0253-7613.93842>

[9] MedDRA Medical Dictionary for Regulatory Activities website accessed in Sept 2018: <https://www.meddra.org/about-meddra/history>

[10] WHO website accessed in Sept 2018: <http://www.who.int/classifications/icd/en/>

[11] PhD thesis, Iain Edward Buchan accessed in sept 2018: <https://statsdirect.co.uk/thesis/md.pdf>

[12] John M. Chambers and Barbara F. Ryan, The ASA Statistical Computing Section: A History The American Statistician, May 1990 (Volume 4, No. 2, pp 87-89).

[13] Central Drugs Standard Control Organization, Government of India Drugs and Cosmetic Act http://www.cdsco.nic.in/writereaddata/Drugs&CosmeticAct.pdf, page 32, (accessed 17-Sept-2017)

[14] Center for Development of Advanced Computing, India, https://cdac.in/index.aspx?id=hi\_dss\_ayusoft, (accessed 17-Sept-2017)

[15] Traditional Knowledge Digital Library, Council of Scientific & Industrial Research (CSIR) and Department of Ayurveda, Yoga & Naturopathy, Unani, Siddha and Homeopathy (AYUSH), Government of India, http://www.tkdl.res.in/tkdl/langdefault/common/Home.asp?GL=Eng, (accessed 17-Sept-2017)

[16] Planning Commission, Government of India, a report by working group on "Access to Health systems including AYUSH" 2006 planningcommission.nic.in/aboutus/committee/wrkgrp11/wg11\_hayush.doc, (accessed 17-Sept-2018)

[17] World Health Organization, a report on WHO traditional medicine strategy: 2014-2023 www.who.int/medicines/publications/traditional/trm\_strategy14\_23/en/, (accessed on 17-Sept-2018)

[18] Planning Commission, Government of India, a report on "12th 5-year plan (2012-2017)" 24th June 2011 planningcommission.gov.in/aboutus/committee/wrkgrp12/health/WG\_7\_ayush.pdf, (accessed on 17-Sept-2018)

[19] Narayana, D. B. A. (2016). Evidence for Ayurvedic products’ efficacy: The devil is in details. Ancient Science of Life, 35(4), 193–194. <http://doi.org/10.4103/0257-7941.188176>

[20] Sarmukaddam, S., Chopra, A., & Tillu, G. (2010). Efficacy and safety of Ayurvedic medicines: Recommending equivalence trial design and proposing safety index. International Journal of Ayurveda Research, 1(3), 175–180. <http://doi.org/10.4103/0974-7788.72491>

[21] S. Rastogi, F. Chiappelli, M.H. Ramchandani, R.H. Singh, Evidence-Based Practice in Complementary and Alterantive Medicine Perspectives, Protocols, Problems and Potential in Ayurveda, 2012, XIV, 245 p.

[22] TDU university website <http://tdu.edu.in/>

[23] PostgreSQL: The World's Most Advanced open source relational database, <https://www.postgresql.org/>

[24] The R Project for Statistical Computing <https://www.r-project.org/>

[25] Tableau software website <https://www.tableau.com/>

[26] Tableau public website <https://public.tableau.com/en-us/s/>

Fig 1: Generic algorithm and steps followed

Use R program to generate tabular or graphical analysis

Use R program to create analysis data tables

Use source tables from the SQLserver

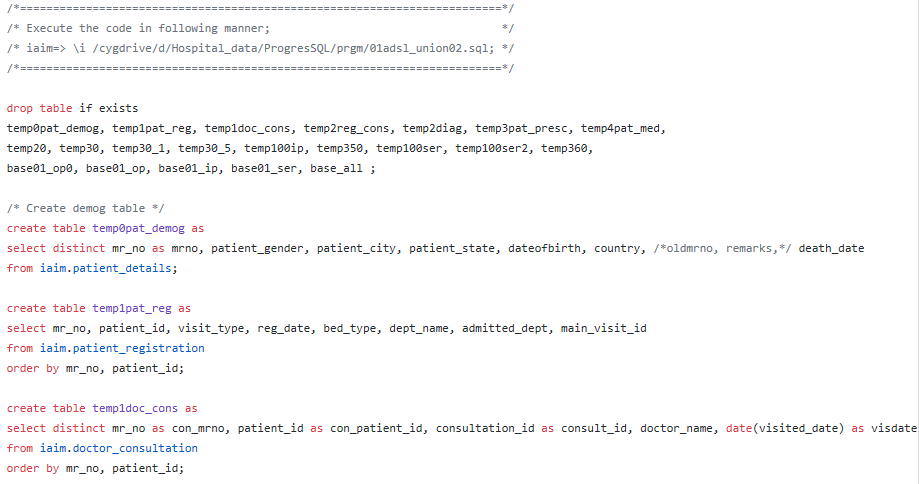
Use SQL queries to combine necessary tables

Use Tableau to generate interactive visual analysis

Figure created by the authors

1. Login on the SQL server using the credentials
2. Use the IAIM schema and access the following tables to generate base table with Demographics information, Patient Visit information and prescribed treatments
   1. STATE\_MASTER
   2. COUNTRY\_MASTER
   3. CITY
   4. STATE
   5. PATIENT\_DETAILS
   6. PATIENT\_REGISTRATION
   7. MRD\_DIAGNOSIS
   8. PATIENT\_PRESCRIPTION
   9. PATIENT\_MEDICINE\_PRESCRIPTIONS
   10. IP\_PRESCRIPTION
   11. SERVICES\_PRESCRIBED
   12. SERVICES
   13. MEDICINE\_SALES\_VIEW
3. There are many CRF pages built to collect relevant Ayurvedic data, measurement data, Hospital visit data, food / exercise advice, etc. This data is present in the following tables:
   1. PATIENT\_SECTION\_DETAILS
   2. PATIENT\_SECTION\_VALUES
   3. SECTION\_MASTER
   4. SECTION\_FIELD\_OPTIONS
   5. SECTION\_FIELD\_DESC
   6. PATIENT\_CONSULTATION\_FIELD\_VALUES
4. The datasets created in steps 2 and 3 are further processed using R programming language and the analysis ready datasets are created

Fig 2: SQL code to extract data from the SQL database



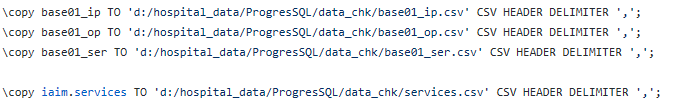


Fig 3: R code, use the input files created using SQL queries



...

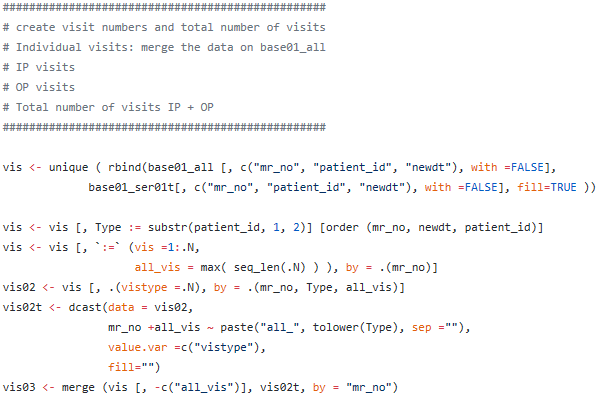
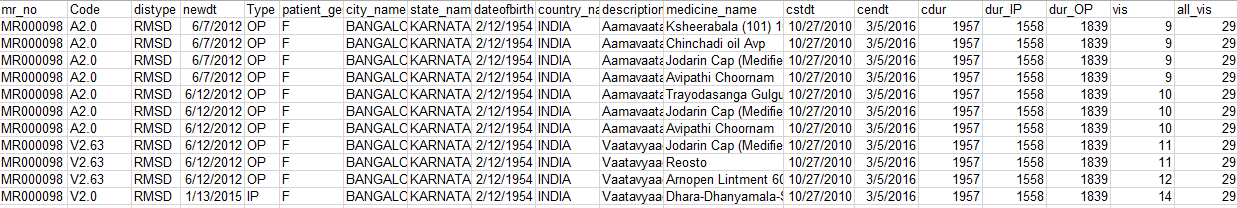


Fig 4: One of the datasets with key variables



The snapshot shows following variables: patient ID, Disease code, Disease type, Date of visit, visit type (IP, OP), gender, city, state, date of birth, Country, Description of disease, Medicine name, Start date, End date, Total duration, In Patient duration, Outpatient duration, Visit number, Total number of visits, etc.

Fig 5: Text table generated using Tableau

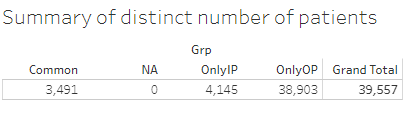


Fig 6: Country wise split, interactive tooltip showing count of unique patients

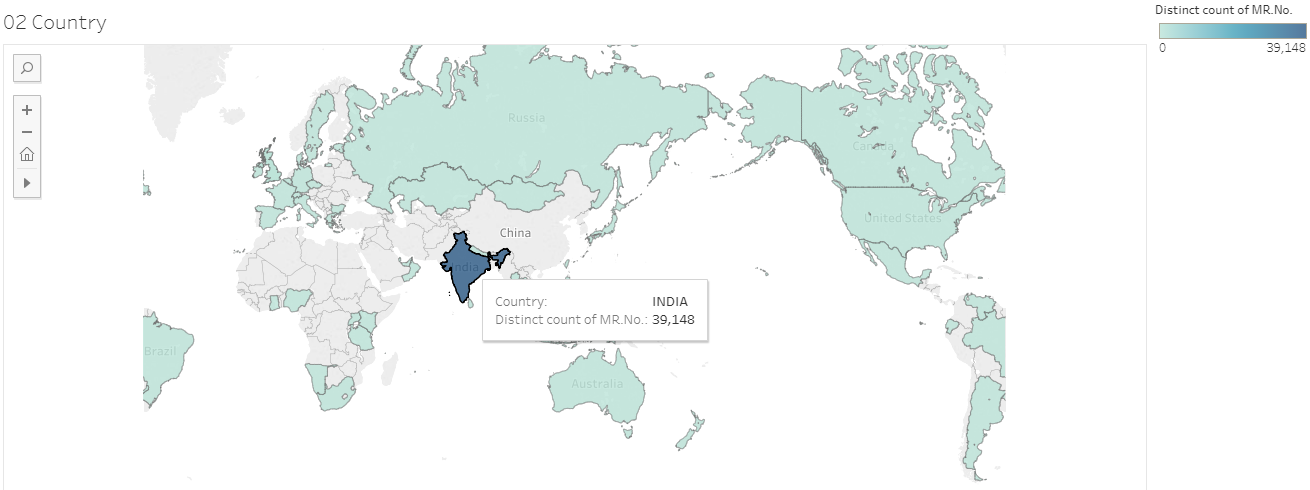


Fig 7: Boxplot for age by country with ability to display or hide data for specific country (is), additional statistics are shown in the tooltip

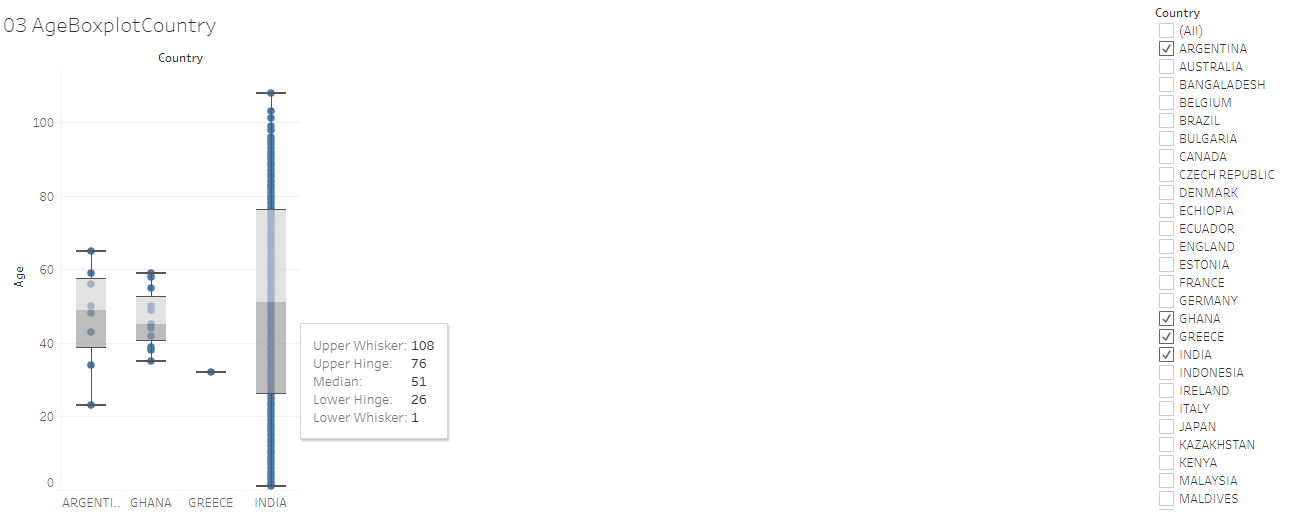


Fig 8: Descriptive statistics table generated using Tableau

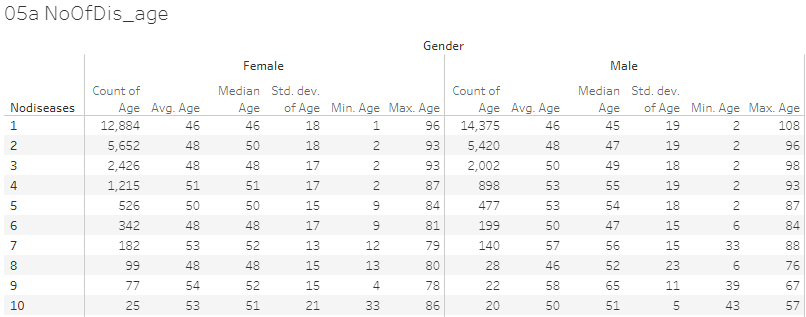


Fig 9: Operational view of number of patients visiting on each day, details of visits displayed in the tooltip maximizing the usage of space

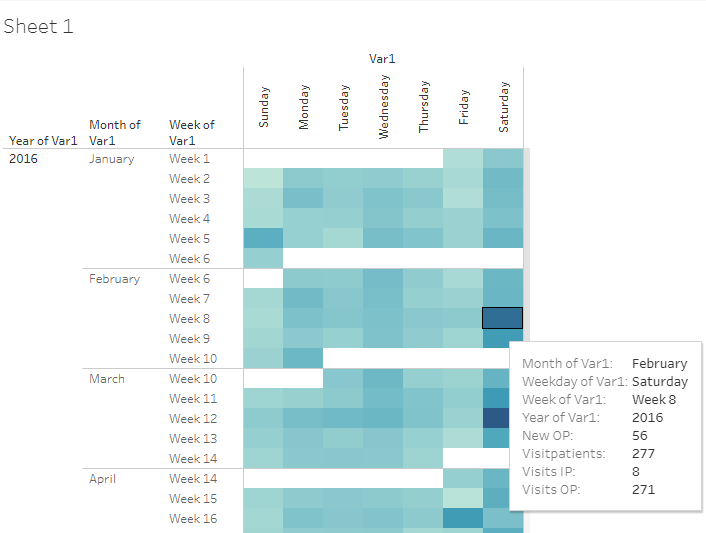
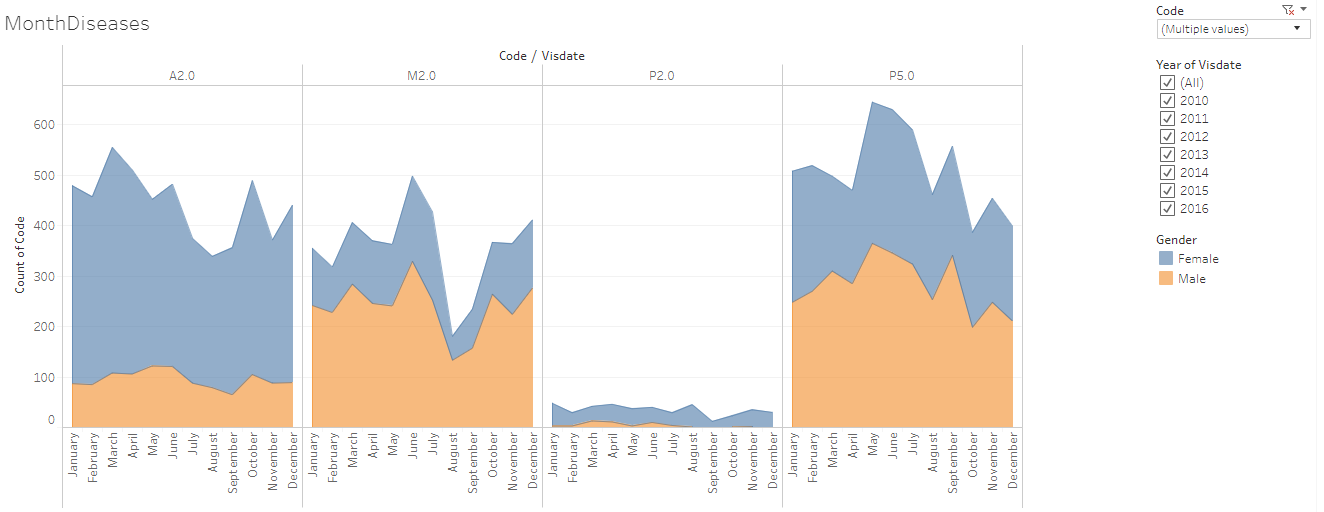
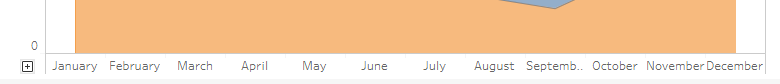


Fig 10: Frequency count of patients for each disease by month by gender. This visual opens for each day, by clicking on “+” sign on the x-axis, showing the operational as well as scientific usefulness.





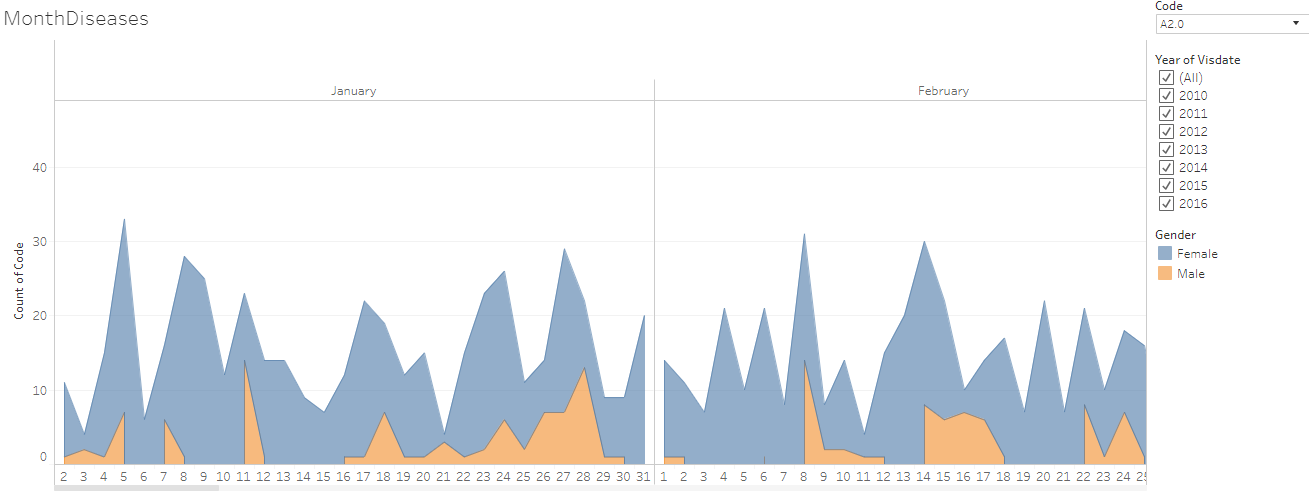


Fig 11: Types of assessment captured by CRFs for individual patients by visit day

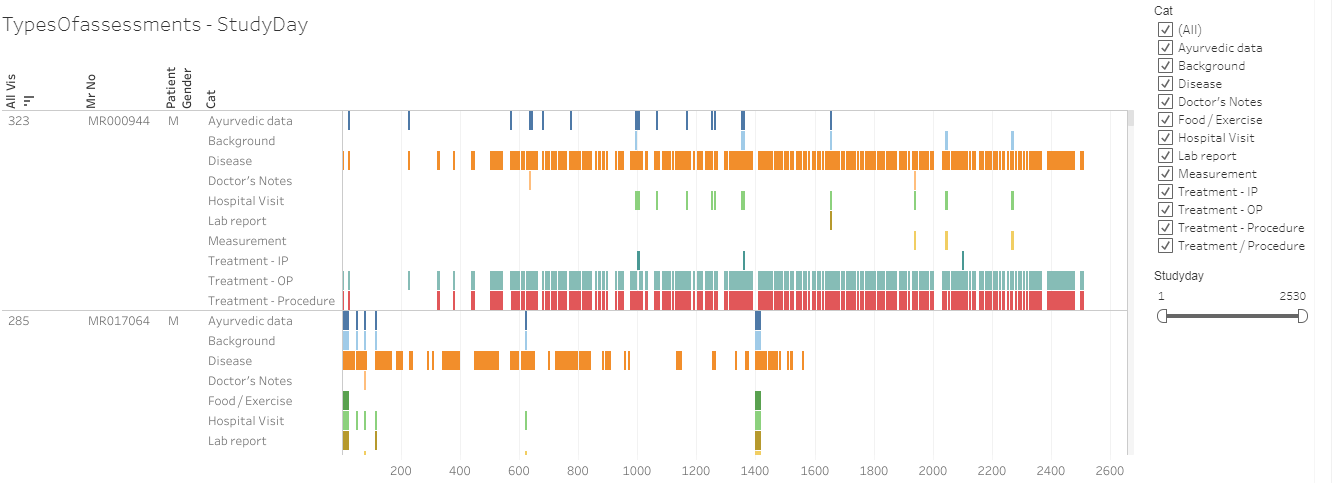


Fig 12: Patient profile view of individual patient by disease and medicine

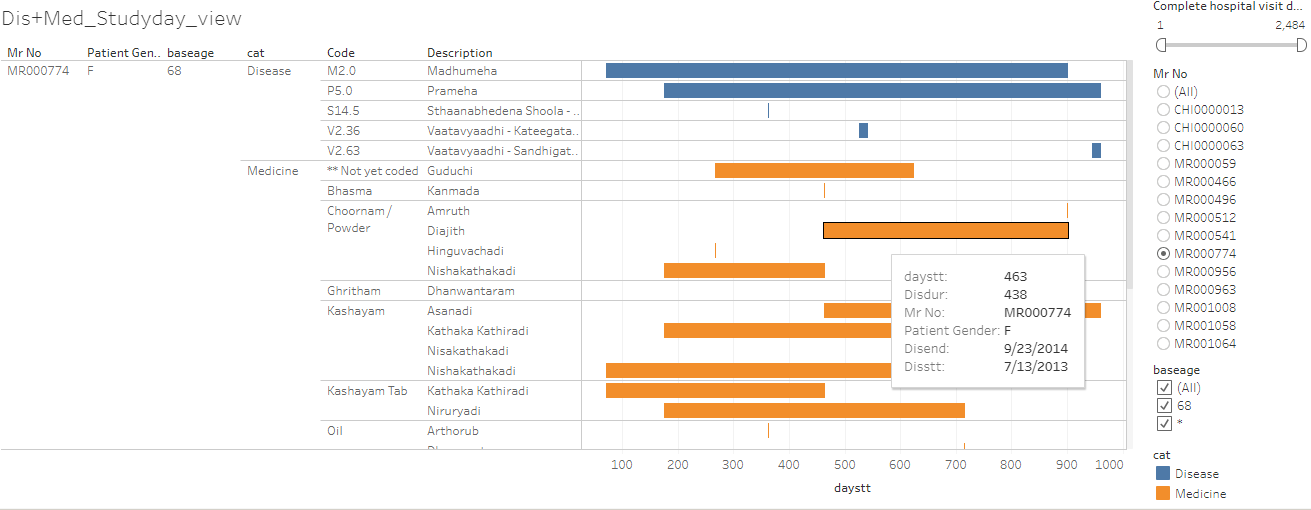


Fig 13: Number of reported diseases and prescribed medicines for each patient by visit

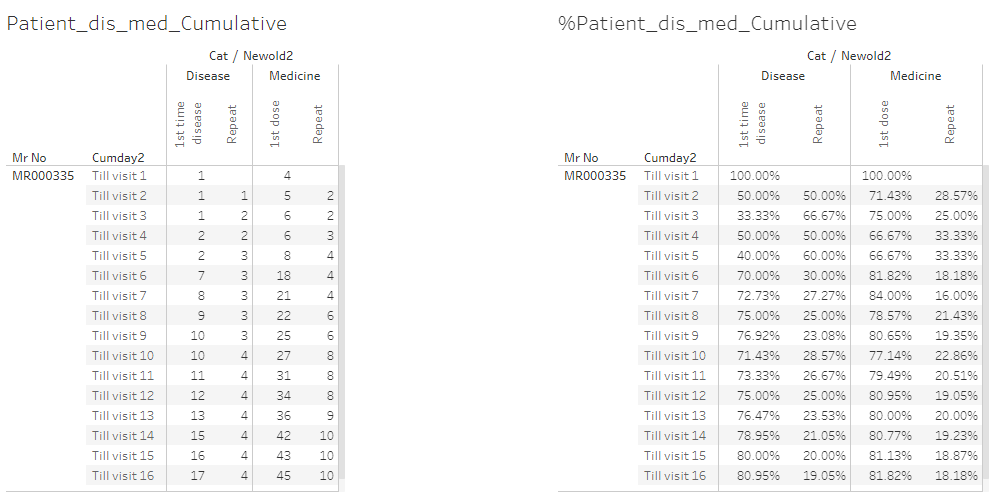


Fig 14: Difference between 2 visits in days for each disease

