### A REPORT

**ON**

**DEVELOPMENT OF AN ALGORITHM TO IDENTIFY INDICATIONS BASED ON TRANSACTION-LEVEL DATA**

#### BY

Sandeep Venkata Kollipara

2013B1A10916G

#### AT



**IQVIA, Inc.**

**A Practice School – II station of**



**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI**

**20th March, 2019**

**A REPORT**

**ON**

**DEVELOPMENT OF AN ALGORITHM TO IDENTIFY INDICATIONS BASED ON TRANSACTION-LEVEL DATA**

**BY**

Sandeep V. Kollipara2013B1A10916GM.Sc.Bio+B.E.Chemical

**Prepared in partial fulfillment of the**

**Practice School – II Course**

**BITS F412**

**AT**

****

**IQVIA, Inc.**

**A Practice School – II station of**



**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI**

**20th March, 2019**

**ACKNOWLEDGEMENT**

I wish to express my gratitude to my mentors Mr. Vijay Viswanathan, Senior Consultant and Ms. Mallika Kowshik, Associate Consultant at Real World Analytics for assigning a project related to my discipline and assisting in getting acclimated to the realm of Analytics in Human Data science.

I would like to extend my thanks to Ms. Karanpreet Kaur and Ms. Faizia Arsheen of the HR department for appointing me as an intern at Commercial Analytics.

I would also like to thank my Practice School – II Instructor Dr. R. Bharathi for supporting me through this wonderful opportunity of a project. I’d like to end by sending my warmest regards to all the colleagues at Commercial Analytics of IQVIA, Inc, Bangalore.

**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE**

**PILANI (RAJASTHAN)**

**Practice School Division**

Station: IQVIA, Inc. Centre: Bangalore

**Duration:** 23rd January, 2019– 18th June, 2019 **Date of Start:** 23rd January, 2019

**Date of Submission:** 20th March, 2018

**Title of the Project:** Development of an algorithm to identify indications based on transaction-level data

**ID No./Name(s)/** 2013B1A10916G/

**Discipline(s)/of** Sandeep V. Kollipara/

**the student(s):** M.Sc.Bio+B.E.Chemical

**Name(s) and** Vijay Viswanathan, Senior Consultant

**Designation(s)** Mallika Kowshik, Assoc Consultant

**of the expert(s):**

**Name(s) of the**

**PS Faculty:** Dr. R. Bharathi, Lecturer, BITS Pilani

**Key Words:** Automated, Python, Transaction-level, Analytics

**Project Areas:** Commercial Analytics

**Abstract:**

This project utilizes analytics on the lucrative biologic market for autoimmune diseases. The prime objective is to develop an open source program on an algorithm to identify indications based on transaction-level data with the functionality to update with the evolving biologic market and requiring no manual intervention from the analyst. The program is built on Python and utilizes an algorithm structured from a model written in SAS and adapts around including additional features that replace the human involvement in the analysis process. The algorithm is divided into 4 phases based on different levels of analysis, sub-objectives and complexity. Post development, the datasets generated by the new program are validated against the pre-existing SAS datasets. The automation aspect is realized by allowing the user to update specs and criteria beforehand instead of hardcoding them within the script. The final dataset obtained has six additional columns per prescription with the details of doctor specialization, indication, therapy duration etc. This dataset is then scheduled to be run in IQVIA’s cockpit software made ready for client use. Being open source, the program reduces the investment on licensed software by the company while suffering no drawbacks in performance. This program can be used for developing programs for other markets as well.

**Signature(s) of Student(s) Signature of PS Faculty**

**Date Date**

**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE**

**PILANI (RAJASTHAN)**

**Practice School Division**

***Response Option Sheet***

**Station:** Beckman Coulter India Pvt. Ltd. **Centre:** Bangalore

**ID. No. & Name(s):** 2013B1A10916G & Sandeep Venkata Kollipara

**Title of the Project:** Development of an algorithm to identify indications based on transaction-level data

Usefulness of the project to the on-campus courses of study in various disciplines. Project should be scrutinized keeping in view the following response options. Write Course No. and Course Name against the option under which the project comes.

Refer Bulletin for Course No. and course Name.

|  |  |  |
| --- | --- | --- |
| **Code**  **No.** | **Response Options** | **Course No. (s) & Name** |
| 1. | A new course can be designed out of this project | No |
| 2. | The Project can help modification of the course content of  Some of the existing Courses | No |
| 3. | The Project can be used directly in some of the existing Compulsory Discipline Courses (CDC)/Discipline Courses Other than Compulsory (DCOC)/ Emerging Area (EA) etc. courses | No |
| 4. | The Project can be used in Preparatory courses like Analysis and Application Oriented Courses (AAOC)/Engineering Science (ES)/Technical Art (TA) and Core Courses | No |
| 5. | This Project cannot come under any of the above mentioned options as it relates to the professional work of the host organization | Yes |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

###### Signature of Student Signature of Faculty

Contents:

|  |  |  |
| --- | --- | --- |
|  | Topic Name | Page No. |
| i | Acknowledgement | 1 |
| ii | Abstract Sheet | 2 |
| iii | Course Sheet | 4 |
| iv | Contents | 5 |
| 1. | Organization | 6 |
| 2. | Project Plan | 8 |
| 3. | Literature Survey | 9 |
| 3.1 | Biologic Market and Autoimmune Diseases | 9 |
| 3.2 | Common Autoimmune Diseases | 11 |
| 4 | Technology and Tools | 12 |
| 5 | Research Methodology | 12 |
| 5.1 | Algorithm | 12 |
| 5.2 | Data Analysis | 14 |
| 5.2.1 | Pandas Framework | 15 |
| 5.2.2 | PYODBC package | 15 |
| 6 | Results and Discussion | 15 |
| 7 | Conclusion | 16 |
| 8 | Future Work | 17 |
| 8.1 | Challenges faced | 17 |
| 9 | Bibliography | 18 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

1. Organization

IMS Health is an American company founded in 1954 by Bill Frohlich and David Dubow. It went private under TPG Capital, CPP Investment Board and Leonard Green & Partners. In 2014, the company went public and started trading in NYSE under the symbol IMS. IMS Health underwent a $17.6 billion merger with Quintiles in 2016 to form QuintilesIMS. In 2017, the company changed its name to IQVIA and changed its symbol to Q in NYSE.

IQVIA is the leading provider in global information and technology services with clients in the Healthcare industry and providing comprehensive solutions to measure and improve their performance. At IQVIA, human data science is applied:

1. by leveraging the analytic rigor and clarity of data science to the ever-expanding scope of human science
2. to enable companies to reimagine and develop new approaches to clinical development and commercialization
3. speed innovation and
4. accelerate improvements in healthcare outcomes.

IQVIA has one of the largest and most comprehensive collections of healthcare information in the world, which includes more than 530 million comprehensive, longitudinal, non-identified patient records spanning sales, prescription and promotional data, medical claims, electronic medical records and social media. Their scaled and growing data set contains approximately 30 petabytes of proprietary data sourced from more than 120,000 data suppliers and covering over 900,000 data feeds globally.

*“We are IQVIA™. Our vision is to outpace the inevitable progress of change across the life sciences and accelerate our ability to empower healthcare decision makers to meet the future head on. We bring the future to clients through Human Data Science. We provide solutions that help our clients innovate with confidence, maximize opportunities and, ultimately, drive healthcare forward.”*

IQVIA, based on the data, delivers information and insights on 85% of the world’s pharmaceuticals, as measured by sales revenue in 2016. IQVIA standardizes, organizes, structures and integrates this data by applying their sophisticated analytics and leveraging its global technology infrastructure to help their clients run their organizations more efficiently and make better decisions to improve their operational and financial performance.

With more than 55,000 employees including approximately 19,000 Commercial Services employees, approximately 29,000 Research & Development Solutions employees and approximately 7,000 Integrated Engagement Services employees, IQVIA conducts operations in more than 100 countries.

**Organization Structure**

Hierarchy:

Chairman/CEO

President (Real World & Analytics Solutions)

Senior Vice President/General Manager

Senior Director/General Manager

Engagement Manager (Advanced Analytics)

Senior Consultant/Associate Consultant

Business Divisions/Domains:

1. Business Integration
2. Centers of Excellence
3. Chief Information Office
4. Chief Medical and Scientific Office
5. Ethics and Compliance Office
6. Finance
7. Legal Office of General Counsel
8. **Global Services**
   1. **Commercial Analytics**
   2. Commercial Outsourcing Solutions
   3. Consulting Services
   4. Primary Intelligence
9. Go-To Market, Sales and Customer Relations
10. Information and Technology Client Solutions
11. Insurance Certificates (Business)
12. Biotech
13. Contract Sales and Medical Solutions
14. Quality Assurance
15. **Real World and Analytics Solutions**
16. Research and Development Solutions
17. Strategy, Marketing and Communications
18. Business Operations
19. Global Environment Health and Safety
    * Business Integration
    * [Centers of Excellence](https://quintiles.sharepoint.com/sites/IQ/BUs/CoE)
    * [Customer CoE](https://quintiles.sharepoint.com/sites/IQ/BUs/CoE/Customer_CoE)
    * [Therapeutic & Speciality CoE](https://quintiles.sharepoint.com/sites/IQ/BUs/CoE/TA_CoE)
    * [Chief Information Office](https://quintiles.sharepoint.com/sites/IQ/BUs/CIO)
    * [Enterprise Systems & Processes](https://quintiles.sharepoint.com/sites/IQ/BUs/CIO/ESandP)
    * [Global IT Service Desk](https://quintiles.sharepoint.com/sites/IQ/BUs/CIO/GlobalITServiceDesk)
    * [Information Security](https://quintiles.sharepoint.com/sites/IQ/BUs/CIO/TechSecurity)
    * [IT Architecture & Strategy](https://quintiles.sharepoint.com/sites/IQ/BUs/CIO/Architect_Strat)
    * [Global Business Continuity](https://quintiles.sharepoint.com/sites/IQ/BUs/CIO/BC)
    * [Enterprise Logging](https://quintiles.sharepoint.com/sites/IQ/BUs/CIO/Enterprise-Logging" \t "_blank)
    * [Chief Medical & Scientific Office](https://quintiles.sharepoint.com/sites/IQ/BUs/ChiefMedSci)
    * [Ethics & Compliance Office](https://quintiles.sharepoint.com/sites/IQ/BUs/ECO)
    * [Buiness Ethics Office](https://quintiles.sharepoint.com/sites/IQ/BUs/ECO/BEO)
    * [Compliance Training Program](https://quintiles.sharepoint.com/sites/IQ/BUs/ECO/ComplianceTraining)
    * [Ethical Moments](https://quintiles.sharepoint.com/sites/IQ/BUs/ECO/EthicalMoments)
    * [Lighthouse](https://quintiles.sharepoint.com/sites/IQ/BUs/ECO/Lighthouse)
    * [Finance](https://quintiles.sharepoint.com/sites/IQ/BUs/Finance)
    * [Legacy IMS Procurement](https://quintiles.sharepoint.com/sites/IQ/BUs/Finance/Temp_Procurement)
    * [Legacy IMS Shared Business Services](https://quintiles.sharepoint.com/sites/IQ/BUs/Finance/FinSBS)
    * [Legacy Q Finance Support Services](https://quintiles.sharepoint.com/sites/IQ/BUs/Finance/FinPSC)
    * [Legacy Q Procurement](https://quintiles.sharepoint.com/sites/IQ/BUs/Finance/Procurement" \t "_blank)
    * [NA Credit & Collections](https://quintiles.sharepoint.com/sites/IQ/BUs/Finance/NA-Credit-And-Collections)
    * [Real Estate & Facilities Management](https://quintiles.sharepoint.com/sites/IQ/BUs/Finance/GREF)
    * [Legal - Office of General Counsel](https://quintiles.sharepoint.com/sites/IQ/BUs/Legal)
    * [Asia Pacific Legal Services](https://quintiles.sharepoint.com/sites/IQ/BUs/Legal/AP)
    * [Contracting](https://quintiles.sharepoint.com/sites/IQ/BUs/Legal/Contracts)
    * [Government Affairs](https://quintiles.sharepoint.com/sites/IQ/BUs/Legal/Government-Affairs)
    * [Japan Legal Services](https://quintiles.sharepoint.com/sites/IQ/BUs/Legal/JP)
    * [Records Management & Retention](https://quintiles.sharepoint.com/sites/IQ/BUs/Legal/RecordsManagement)
    * [Securities](https://quintiles.sharepoint.com/sites/IQ/BUs/Legal/Quintiles-Securities)
    * [Intellectual Property and Technology](https://quintiles.sharepoint.com/sites/IQ/BUs/Legal/IP)
    * [Third Party Access](https://quintiles.sharepoint.com/sites/IQ/BUs/Legal/Third-Party-Access" \t "_blank)
    * [Global Services](https://quintiles.sharepoint.com/sites/IQ/BUs/GlServices_Consult)
    * [Commercial Analytics](https://quintiles.sharepoint.com/sites/IQ/BUs/GlServices_Consult/ACA)
    * [Commercial Outsourcing Services](https://quintiles.sharepoint.com/sites/IQ/BUs/GlServices_Consult/COS)
    * [Consulting Services](https://quintiles.sharepoint.com/sites/IQ/BUs/GlServices_Consult/CS)
    * [Primary Intelligence](https://quintiles.sharepoint.com/sites/IQ/BUs/GlServices_Consult/PI)
    * [Go-to-Market, Sales & Customer Relations](https://quintiles.sharepoint.com/sites/IQ/BUs/GTM)
    * [Commercial Business Office](https://quintiles.sharepoint.com/sites/IQ/BUs/GTM/Enterprise%20Solutions%20Development)
    * [Emerging Biopharma (EBP)](https://quintiles.sharepoint.com/sites/IQ/BUs/GTM/EliteAccounts)
    * [Global Business Operations](https://quintiles.sharepoint.com/sites/IQ/BUs/GTM/GlBusOpns)
    * [National & Sub-National Sales Management](https://quintiles.sharepoint.com/sites/IQ/BUs/GTM/SalesMgmt)
    * [R&DS Sales & Account Management](https://quintiles.sharepoint.com/sites/IQ/BUs/GTM/CSMG" \t "_blank)
    * [Global Sales Excellence](https://quintiles.sharepoint.com/sites/IQ/BUs/GTM/GlSalesExc)
    * [Information & Technology Client Solutions](https://quintiles.sharepoint.com/sites/IQ/BUs/InfoTechSol_Client)
    * [Consumer Health](https://quintiles.sharepoint.com/sites/IQ/BUs/InfoTechSol_Client/ConsumerHlth)
    * [Global Market Insights](https://quintiles.sharepoint.com/sites/IQ/BUs/InfoTechSol_Client/Gl_Mktg_Insght)
    * [Pharmacy Services](https://quintiles.sharepoint.com/sites/IQ/BUs/InfoTechSol_Client/Pharmacy-Services)
    * [Technology Solutions](https://quintiles.sharepoint.com/sites/IQ/BUs/InfoTechSol_Client/TechSols)
    * [Insurance Certificates (Business)](https://quintiles.sharepoint.com/sites/IQ/BUs/RiskManagement)
    * [Certificates & Evidence of Insurance](https://quintiles.sharepoint.com/sites/IQ/BUs/RiskManagement/InsuranceCertificates)
    * [Contractual Insurance Language & Guidelines](https://quintiles.sharepoint.com/sites/IQ/BUs/RiskManagement/InsuranceGuidelines)
    * [Contract Sales and Medical Solutions](https://quintiles.sharepoint.com/sites/IQ/BUs/IES)
    * [CSMS Europe](https://quintiles.sharepoint.com/sites/IQ/BUs/IES/IES-Europe)
    * [CSMS North America](https://quintiles.sharepoint.com/sites/IQ/BUs/IES/IES-North-America)
    * [Quality Assurance](https://quintiles.sharepoint.com/sites/IQ/BUs/QA_Glocal)
    * [Commercial QA](https://quintiles.sharepoint.com/sites/IQ/BUs/QA_Glocal/Comml)
    * [EDA](https://quintiles.sharepoint.com/sites/IQ/BUs/QA_Glocal/EDA)
    * [R & D QA](https://quintiles.sharepoint.com/sites/IQ/BUs/QA_Glocal/RandD_QA)
    * [RWI QA](https://quintiles.sharepoint.com/sites/IQ/BUs/QA_Glocal/RWI_QA)
    * [Systems Compliance Office (SCO) QA](https://quintiles.sharepoint.com/sites/IQ/BUs/QA_Glocal/SCO)
    * [Business Analytics & Decision Support (BA&DS)](https://quintiles.sharepoint.com/sites/IQ/BUs/QA_Glocal/BA&DS)
    * [Business Support Services](https://quintiles.sharepoint.com/sites/IQ/BUs/QA_Glocal/BSS)
    * [Real-World & Analytics Solutions](https://quintiles.sharepoint.com/sites/IQ/BUs/RWI)
    * [Research & Development Solutions](https://quintiles.sharepoint.com/sites/IQ/BUs/RandDSol)
    * [Clinical Operations](https://quintiles.sharepoint.com/sites/IQ/BUs/RandDSol/ClinOps)
    * [Data Sciences, Safety & Regulatory](https://quintiles.sharepoint.com/sites/IQ/BUs/RandDSol/DSSR)
    * [Early Clinical Development](https://quintiles.sharepoint.com/sites/IQ/BUs/RandDSol/PhaseI)
    * [Global Functional Resourcing](https://quintiles.sharepoint.com/sites/IQ/BUs/RandDSol/GFR)
    * [Libraries & Information Services Center](https://quintiles.sharepoint.com/sites/IQ/BUs/RandDSol/Libraries)
    * [Q2 Solutions](https://quintiles.sharepoint.com/sites/IQ/BUs/RandDSol/JVLabs)
    * [Research & Development Technology](https://quintiles.sharepoint.com/sites/IQ/BUs/RandDSol/RDS_Tech)
    * [Strategy, Marketing & Communications](https://quintiles.sharepoint.com/sites/IQ/BUs/SMC)
    * [Enterprise Digital](https://quintiles.sharepoint.com/sites/IQ/BUs/SMC/DS)
    * [Media Relations & Communications](https://quintiles.sharepoint.com/sites/IQ/BUs/SMC/CorpComm)
    * [IQVIA Institute](https://quintiles.sharepoint.com/sites/IQ/BUs/SMC/Institute)
    * [Creative & Marketing Operations](https://quintiles.sharepoint.com/sites/IQ/BUs/SMC/CCOE)
    * [Customer Experience & Insights](https://quintiles.sharepoint.com/sites/IQ/BUs/SMC/CEI)
    * [Global Marketing](https://quintiles.sharepoint.com/sites/IQ/BUs/SMC/Global%20Marketing)
    * [Sales Excellence](https://quintiles.sharepoint.com/sites/IQ/BUs/SMC/Sales_Excellence)
    * [Strategic Planning](https://quintiles.sharepoint.com/sites/IQ/BUs/SMC/Strategic_Planning)
    * [Business Operations](https://quintiles.sharepoint.com/sites/IQ/BUs/Business_Operations)
    * [Global Environmental Health & Safety](https://quintiles.sharepoint.com/sites/IQ/BUs/EHS)

Important Personnel:

* Chairman & CEO: Ari Bousbib
* President (Real World Analytics & Solutions): Jon Resnick
* Senior Vice President & General Manager: Prashant Parab
* Senior Director & General Manager: Jaivardhan Iyer
* Engagement Manager(Advanced Analytics): Sunil Kumar Singh

1. Project Plan:

Figure 1: Gannt Chart of the project timeline with blue denoting completed and yellow denoting currently underway

The project spans 6 weeks from 4th Feb to 20th March with the first week spent training on Python and learning the data structures.

The aim of the project is to develop an algorithm to identify indication based on transaction-level data with robustness to update.

1. Prescription-level Assignment of Doctor Specialization
2. Patient-level Assignment of Doctor Specialization and Classification
3. Indication Split: Dermatology Patients into Hidradenitis Suppurative (HS) and Psoriasis (PSO).
4. Therapy Duration Calculation per dosage unit and Assignment.

The initial 2 phases require fundamental understanding of the dataset and importing them into python environment to access, search and modify for analysis. The SAS model must be deciphered alongside development of the code especially in the 3rd Phase along with the implementation of SQL access for the program. The final phase involves calculating the therapy duration spec based on multiple criteria from different levels of the dataset.

Automation of the program enters the 2nd and 4th phases which employs classifying data by specific criterion which change with time depending on the markets.

1. Literature survey

3.1 Biologic Market and Autoimmune Diseases:

The market in focus is of Biologic Therapies for Autoimmune diseases. A Biologic is a product extracted or semi-synthesized from living organism most of them manufactured utilizing Recombinant DNA Technology. Their structure is not characterized and are significantly larger in size and complicated compared to drugs. Delivery systems are primarily in parenteral route, they are injected or infused directly in the target region or released into the vein.

“The process is the product” aptly emphasizes the quality control restraints and effectiveness of the molecule as manufacturing is concerned. Biologics market requires heavy nascent investments. They also impact economically, for e.g., Rheumatoid Arthritis, Drug therapy like methotrexate of cost is less than $100 and its counterpart biologics certolizumab and abatacept costs $300-6000.

The pros brought by biologics are:

* High selectivity in action,
* potent therapeutic efficiency and
* limited side effects.

The risks/drawbacks entailing biologics are:

* High cost,
* long term use increasing risk of cancer and
* less experience in clinical field

Biologics being a lucrative market and Autoimmune diseases being lifestyle-affecting to fatal, the prescriptions and injections are taken strictly on time thereby making this market analysis most accurate.

The biologics for autoimmune diseases are broadly classified into 4 categories which are explained as follows:

TNF Inhibitors:

Examples of these drugs certolizumab (*Cimzia*), etanercept, golimumab (*Simponi*), adalimumab (*Humira*) and infliximab (*Remicade*). The reduce inflammation and can be used in combinations of 2-3 doses. All of them have been clinically approved for children and certolizumab also for pregnant women.

IL Inhibitors:

Examples include anakinra (*Kineret*)[IL-1 inhibitor], tocilizumab (*Actemra*)[IL-6 inhibitor], canakinumab (*Ilaris*), secukinumab (*Cosentyx*) and ustekinumab (*Stelara*). They are used after TNF inhibitors known to be effective and safe. Rarely bowel perforations are seen in some cases.

B-cell Inhibitors

Examples include belimumab (*Benlysta*) and rituximab (*Rituxan*). They interfere with production of abnormal antibodies (produced from B-cells. About 2 infusions are conducted in a year with relative long-term safety. Risks include blood pressure changes, chest pain, difficulty breathing, rash, dizziness and/or flu-like symptoms which need additional control medications.

T-cell Inhibitors

Some of them are also known as Selective Co-stimulation Modulators with examples like Abatacept (*Orencia*). Effects are not seen until 4-6 weeks after treatment and is more effective when used in combination with other common drugs. Risks involve being susceptible to infections such as pneumonia, tuberculosis and influenza.

3.2 Common Autoimmune Diseases:

Rheumatoid Arthritis: Inflammation of Joint Linings. All TNF inhibitors and adalimumab-atto (*Amjevita*) a biosimilar to Humira can be prescribed.

Multiple Sclerosis: Immune system attacks myelin sheath and deteriorates nerves with potential permanent damage. It affects limbs and the root cause is unknown. Biologic natalizumab (*Tysabri*) by Novartis in the market with Avonex (*Interferon B*) & Capaxone (*glatiramer acetate*) drugs (non-biologics).

Lupus: Immune system attacks tissues and organs and its severity can cause permanent tissue damage. Biologic belimumab (*Benlysta*) is the first approved biologic for Lupus treatment. It can also be used alongside other immunosuppressive drugs (non-biologics.

Psoriasis: Immune reaction on skin resulting in scale formation and discoloration.

Most TNF inhibitors like ustekinumab (*Stelara*) [IL-12/23 inhibitor], secukinumab (*Cosentyx*) and ixekizumab (*Taltz*) [IL-17 inhibitors], guselkumab (*Tremfya*) and tildrakizumab-asmn (*Ilumya*) [T-cell inhibitor].

Type-I Diabetes: Immune system destroys pancreatic beta cells completely leaving the victim unable to secrete insulin for metabolism. Common daily insulin injections are taken by the patients. Insulin by itself is a biological compound albeit regulated as a common drug.

Inflammatory Bowel Disease: Crohn's disease and ulcerative colitis come under this.

Most TNF inhibitors can be prescribed and biologics ustekinumab (only for Crohn’s disease) and vedolizumab (for both diseases) are new.

1. Technology and Tools used:

Software:

* Spyder 3.2, a Python 3.6 Integrated Development Environment (IDE)
* Microsoft Excel 2016, data viewer in spreadsheets

Technology:

* Pandas Framework, to use Data frames and to read, access and modify Excel, SAS datasets.
* PYODBC package, to access database server.

Database:

* Confidential IQVIA datasets of longitudinal transactions, stored both locally and on SQL server.

1. Research Methodology:

With the project being a culmination of healthcare and analytics, there are a plethora of concepts that come into play, the healthcare part requires knowledge of the biologics, indications and the therapies involved to formulate an algorithm whereas the analytics part requires technical knowledge in programming to implement the idea into a program. Hence, the concepts are vastly different for the algorithm and its implementation.

**5.1 Algorithm:**

The 4 phases of the program require selection, classification, spec calculation and spec assignment at prescription (referred to as ‘rx’ in the algorithm), patient, indication and doctor specialty levels.

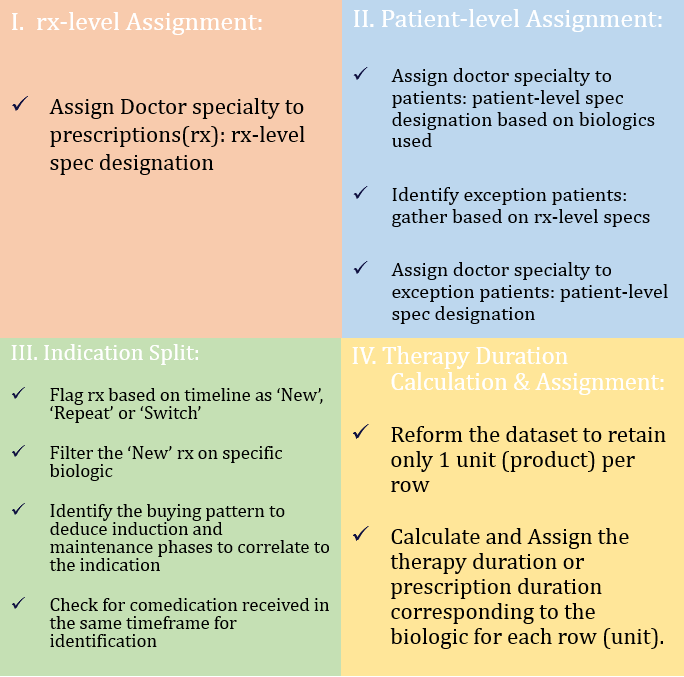


Figure 2: The algorithm of the 4 phases of project depicted above.

The base dataset containing the transaction-level data of patients corresponding to their prescriptions is of only biologics. Each transaction is designated a ‘Doctor Specialty’ spec based on what doctor prescribed it. The spec designated is of either ‘Rheumato’ (Rheumatology), ‘Dermato’ (Dermatology), ‘Gastro’ (Gastroenterology) or ‘Irrelevant’ for any other specialization. The ends Phase I.

The dataset resulting from last phase is grouped by patient ID and the transactions of each patient are checked. If the biologic they are purchasing are exclusively prescribed by a Rheumatologist, Dermatologist or Gastroenterologist, then they’re designated a patient-level spec of ‘Rheumato’, ‘Dermato’ or ‘Gastro’ respectively. The remaining patients without any exclusive purchases are checked for the number of transactions per unit of ‘Rheumato’, ‘Dermato’, ‘Gastro’ or ‘Irrelevant’ spec assigned in Phase I. The patient-level spec is assigned on the basis of the higher number of transaction-level spec. If there is an equality, the precedence order depending on the indication and biologic of interest. This ends Phase II.

The dataset resulting from last phase is considered again and flagged transaction wise ‘New’ if it is the first ever biologic prescribed or one initiated after a gap of more than 18 months. The new patients are filtered on the biologic of interest and checked for comedications that come with the indications and split based on them. By the end, the dataset is split across Hidradenitis Suppurativa or Psoriasis and flagged accordingly as ‘HS’ or ‘PSO’ under Indication column. This concludes Phase III.

The Indication flagged dataset from Phase III is currently in transaction per row format. It is converted to unit (of product) per row format and then numbered pack-wise from their newest transaction to the end or another ‘new’ transaction after a 18 month gap. Then they are assigned therapy duration (days) per unit calculated on the basis of the induction and maintenance phases of the therapy followed by them. This concludes Phase IV.

**5.2 Data Analysis:**

The primary tool used by an analyst in IQVIA is SAS, a licensed software suite with point-and-click graphical interface for non-technical users along with advanced SAS code capabilities. Python is free, open source software which is lightweight and rich in resources in the form of libraries extending its reaches to Analytics making it the best open source alternative to SAS for Data Scientists. This project involves developing a Python program as an alternative to an existing SAS model in Autoimmune Biologics market.

5.2.1 Pandas Framework:

The ‘pandas.DataFrame’ module is used as the primary data structure to encapsulate the datasets of longitudinal transactions in SAS files and SQL server. The ‘pandas.iloc’ function was used to access the dataset index-wise and also modify it. The ‘pandas.concat’ function was used to append data frames post classification by ‘pandas.groupby’. ‘pandas.iterrows’ was used for row-wise access.

5.2.2 PYODBC Package:

The ‘pyodbc’ python package is an Open Database Connectivity (ODBC) tool allows connection to multiple types of SQL based servers such as Oracle, MySQL, Microsoft SQL Server, PostgreSQL, SAP HANA, Sybase ASE and DB2.

1. Results and Discussion





Figure 3 (a) & (b): Data of a patient post analysis by the program.

The final dataset exported to an excel spreadsheet is filtered to a Psoriasis patient’s transactions and shows the additional 6 columns added to it during the analysis. The last 6 columns are:

* DoctorClass: the prescription-level doctor specialty assignment
* PatientLevelClass: the patient-level doctor specialty assignment
* ProdTransaction: the flag describing if prescription is new, repetition or switched depending on patient’s history
* Indication: The flag assigned in 3rd Phase identifying him/her with the indication
* PackNumber: The count of the unit of product used by the patient across history
* RXDuration: The length of duration of the particular unit of prescription depending on induction or maintenance phase.

1. Conclusion

At the time of writing this report, Phase III is still underway, and the results depicted show a retrospectively assigned Indication stat which is to be originally assigned in Phase III. Functionality for updating was implemented which allows user to set criterion for analysis by retrieving data from Excel worksheets in 2nd and 4th phases. The verification of the data is only done for Phase I code. The project objectives are estimated to be accomplished by March 20th but additional features like GUI, summary file creation and update functionality may extend the deadline till March 27th.

The program once completed will require only the base dataset (biologics) and SQL dataset (comedication) after which the analysis is automated completely. With the evolving biologics markets, the program allows for updating specs without the manual intervention of the user. The fully automated program with the means of updating to the current market developments with the rise of new biologics and biosimilar means the analyst not requiring coding or have any technical knowledge to run the analysis on a dataset. Biologics being a lucrative market and Autoimmune diseases being lifestyle-affecting to fatal, the prescriptions and injections are taken strictly on time thereby making this market analysis most accurate.

1. Future Work

Being open source, the program reduces the investment on licensed software by the company while suffering no drawbacks in performance. This model can be used for developing programs for other markets as well.

**8.1 Challenges faced:**

1. SAS Programming language
   * To understand the pre-existing model by studying the code for deducing the algorithm.
2. Learning the structure of transaction-level data
   * Selection of python data structure for optimal analysis among csv (comma separated values), lists and Dataframes.
3. Conversion of datasets from SAS to Python environment
   * With each data structure having its own features and limitations with respect to semantics and warnings.
   * ‘SettingWithCopyWarning’ in Pandas while modifying dataframes.
4. Accessing SQL databases
   * Learning about different SQL servers like MSSQL, PostgreSQL, MySQL etc
   * Accessing the database using python.
5. Update Capability
   * To allow input of user defined criteria for analysis by creating a robust program.

# Bibliography

Arthritis Foundation. (n.d.). *Biologics (Biologic Response Modifier) Overview*. Retrieved from Arthritis Foundation: https://www.arthritis.org/living-with-arthritis/treatments/medication/drug-types/biologics/drug-guide-biologics.php

Chron's & Colitis UK. (2017). *Treatments*. Retrieved from Chron's & Colitis UK: https://www.crohnsandcolitis.org.uk/about-inflammatory-bowel-disease/treatments

Denise Mann. (n.d.). *New Drugs for Rheumatoid Arthritis: Is a Biologic Pill on the Way?* Retrieved from WebMD: https://www.webmd.com/rheumatoid-arthritis/features/new-drugs-for-ra#1

FDA. (n.d.). *U.S. Food & Drug Administration*. Retrieved from Fractionated Plasma Products > PANZYGA: https://www.fda.gov/BiologicsBloodVaccines/BloodBloodProducts/ApprovedProducts/LicensedProductsBLAs/FractionatedPlasmaProducts/ucm615698.htm

FDA. (n.d.). *Vaccines, Blood & Biologics > CUTAQUIG*. Retrieved from U.S. Food & Drug Administration: https://www.fda.gov/BiologicsBloodVaccines/ucm628258.htm

Lindsey, H. (2018, 3 7). *Common Treatments for Lupus*. Retrieved from Everyday Health: https://www.everydayhealth.com/lupus/guide/treatment/

Morriss, E. (2019, 2 4). *Rise of the biosimilars*. Retrieved from Pharmafield: https://pharmafield.co.uk/in\_depth/rise-of-biosimilars/

N., S.-B. (2014). Biologics: the role of delivery systems in improved therapy. *Biologics*, 8:107-14.

Rath, L. (2017, June 1). *FDA OKs a New Biologic for RA*. Retrieved from Arthritis Foundation: http://blog.arthritis.org/news/fda-approves-new-rheumatoid-arthritis-biologic-sarilumab/

Rossi, K. (2018, OCTOBER 31). *Biosimilar, Hyrimoz, Approved by FDA for Host of Chronic Conditions*. Retrieved from MD Magazine: https://www.mdmag.com/medical-news/biosimilar-hyrimoz-approved-fda-chronic-conditions

Shanley, M. (2017, August 29). *FDA Approves Biosimilar for Chronic Inflammatory Diseases*. Retrieved from Rare Disease Report: https://www.raredr.com/news/fda-approves-inflammatory-diseases-biosimilar

Weiss, M. (2017, 12 14). *FSA Approves Biosimilar for Autoimmune Diseases*. Retrieved from The Dermatologist: https://www.the-dermatologist.com/news/fda-approves-biosimilar-autoimmune-diseases