Cytel

# Analysis and Visualization of Pharmacovigilance Data (RWE) Using R

*Kishore Hari*
*Cytel, Pune*

Cytel

# RWE in Healthcare, Pharmacovigilance

Cytel

- Pharmacovigilance - science of detection, assessment and prevention of Adverse Drug Reactions (ADR) post approval of drugs

# RWE in Healthcare, Pharmacovigilance

Cytel

- Pharmacovigilance - science of detection, assessment and prevention of Adverse Drug Reactions (ADR) post approval of drugs
- Multiple organizations collect and handle the data individually

# RWE in Healthcare, Pharmacovigilance

Cytel

- Pharmacovigilance - science of detection, assessment and prevention of Adverse Drug Reactions (ADR) post approval of drugs
- Multiple organizations collect and handle the data individually
- Aim: To develop a tool that can integrate data from multiple sources, analyze and visualise it

# RWE in Healthcare, Pharmacovigilance

Cytel

- Pharmacovigilance - science of detection, assessment and prevention of Adverse Drug Reactions (ADR) post approval of drugs
- Multiple organizations collect and handle the data individually
- Aim: To develop a tool that can integrate data from multiple sources, analyze and visualise it
- Question: What capabilities are needed?

Cytel

# Nature of Pharmacovigilance Data

- Large data collected in over 5 decades. Every year more data pours in.

# Nature of Pharmacovigilance Data

- Large data collected in over 5 decades. Every year more data pours in.
- Two broad categories

# Nature of Pharmacovigilance Data

- Large data collected in over 5 decades. Every year more data pours in.

- Two broad categories
  1. Structured data
     - ★ Spontaneous Reporting Systems (SRS)
     - ★ Well defined data fields and data types
     - ★ Govt. organizations : FDA, WHO, Eudra, Health Canada
     - ★ Private organizations : Pharmaceutical companies

# Nature of Pharmacovigilance Data

- Large data collected in over 5 decades. Every year more data pours in.

- Two broad categories
  1. Structured data
     - ★ Spontaneous Reporting Systems (SRS)
     - ★ Well defined data fields and data types
     - ★ Govt. organizations : FDA, WHO, Eudra, Health Canada
     - ★ Private organizations : Pharmaceutical companies
  2. Unstructured data
     - ★ Social media, medical forums etc.
     - ★ Data generated by/from non-professionals - incomplete and inaccurate data
     - ★ Textual data that has to be mined for drug-adverse event reports

# Nature of Pharmacovigilance Data

- Large data collected in over 5 decades. Every year more data pours in.

- Two broad categories
  1. Structured data
     - ★ Spontaneous Reporting Systems (SRS)
     - ★ Well defined data fields and data types
     - ★ Govt. organizations : FDA, WHO, Eudra, Health Canada
     - ★ Private organizations : Pharmaceutical companies
  2. Unstructured data
     - ★ Social media, medical forums etc.
     - ★ Data generated by/from non-professionals - incomplete and inaccurate data
     - ★ Textual data that has to be mined for drug-adverse event reports

Let's talk about the structured data!

# Formats of Structured Data

# Formats of Structured Data

1. Server-hosted databases: e.g. Eudra, VigiBase
   - Relational database systems, easy to query data
   - Requires maintainance
   - Restricted access

# Formats of Structured Data

Cytel

1. Server-hosted databases: e.g. Eudra, VigiBase
   - Relational database systems, easy to query data
   - Requires maintainance
   - Restricted access

2. Text/XML files: e.g. FAERS and Health Canada; VigibaseExtract queried data

# Formats of Structured Data

Cytel

1. Server-hosted databases: e.g. Eudra, VigiBase
   - Relational database systems, easy to query data
   - Requires maintainance
   - Restricted access

2. Text/XML files: e.g. FAERS and Health Canada; VigibaseExtract queried data
   - Delimited text files that contain a table per file (e.g. drug, demographics etc.)

# Formats of Structured Data

Cytel

1. Server-hosted databases: e.g. Eudra, VigiBase
   - Relational database systems, easy to query data
   - Requires maintainance
   - Restricted access

2. Text/XML files: e.g. FAERS and Health Canada; VigibaseExtract queried data
   - Delimited text files that contain a table per file (e.g. drug, demographics etc.)
   - XML files - each ICSR converted into a node, makes them huge!

# Formats of Structured Data

Cytel

1. Server-hosted databases: e.g. Eudra, VigiBase
   - Relational database systems, easy to query data
   - Requires maintainance
   - Restricted access

2. Text/XML files: e.g. FAERS and Health Canada; VigibaseExtract queried data
   - Delimited text files that contain a table per file (e.g. drug, demographics etc.)
   - XML files - each ICSR converted into a node, makes them huge!
   - Relational database setup is recommended

# Current Methods to Access the Data

# Current Methods to Access the Data

- FAERS and Health Canada
  - Data is open for downloading
  - Web tools to search for information on specific drugs is made available (openFDA)

# Current Methods to Access the Data

- FAERS and Health Canada
  - Data is open for downloading
  - Web tools to search for information on specific drugs is made available (openFDA)
- WHO: The biggest pharmacovigilance program
  - Data is managed by Uppsala Monitoring Center (VigiBase)
  - Public access via VigiAccess - web tool
  - Downloading data requires subscription

# Current Methods to Access the Data

- FAERS and Health Canada
  - Data is open for downloading
  - Web tools to search for information on specific drugs is made available (openFDA)
- WHO: The biggest pharmacovigilance program
  - Data is managed by Uppsala Monitoring Center (VigiBase)
  - Public access via VigiAccess - web tool
  - Downloading data requires subscription
- Eudra : EudraVigilance program
  - Access via EVDAS
  - Requires subscription

# Proposal

Development of a tool which can

# Proposal

Development of a tool which can

- Integrate data from different databases, i.e., has the capability to work with different data types

# Proposal

Development of a tool which can

- Integrate data from different databases, i.e., has the capability to work with different data types
- Able to handle large amounts of data

# Proposal

Development of a tool which can

- Integrate data from different databases, i.e., has the capability to work with different data types
- Able to handle large amounts of data
- Analyse the data using various statistical methods

# Proposal

Development of a tool which can

- Integrate data from different databases, i.e., has the capability to work with different data types
- Able to handle large amounts of data
- Analyse the data using various statistical methods
- Create appropriate visualizations for the data

# Proposal

Development of a tool which can

- Integrate data from different databases, i.e., has the capability to work with different data types
- Able to handle large amounts of data
- Analyse the data using various statistical methods
- Create appropriate visualizations for the data

Ideal situation: deploy it on the web

Cytel

# R Language

- Originally created as a tool for statistical computing

# R Language

- Originally created as a tool for statistical computing
- Equipped with multiple statistical analysis algorithms as easy-to-use functions

# R Language

- Originally created as a tool for statistical computing
- Equipped with multiple statistical analysis algorithms as easy-to-use functions
- Open source, a huge and reliable community

# R Language

- Originally created as a tool for statistical computing
- Equipped with multiple statistical analysis algorithms as easy-to-use functions
- Open source, a huge and reliable community
- Acquired strong tools of data analysis and visualization over the past decade

# Can R Handle Large Data?

# Can R Handle Large Data?

Yes! It can!

# Can R Handle Large Data?

Yes! It can!

- Piping and lazy loading to avoid committing data to memory unless absolutely neccessary (***dplyr***)

# Can R Handle Large Data?

Yes! It can!

- Piping and lazy loading to avoid committing data to memory unless absolutely neccessary (***dplyr***)
- Packages that can create and communicate with relational databases (***sqldf, dplyr***)

# Can R Handle Large Data?

Yes! It can!

- Piping and lazy loading to avoid committing data to memory unless absolutely neccessary (***dplyr***)
- Packages that can create and communicate with relational databases (***sqldf, dplyr***)
- Partial/meta loading of data (***feather***)

# Can R Handle Large Data?

Yes! It can!

- Piping and lazy loading to avoid committing data to memory unless absolutely neccessary (***dplyr***)
- Packages that can create and communicate with relational databases (***sqldf, dplyr***)
- Partial/meta loading of data (***feather***)
- Packages to handle Large data (***big memory, ff*** etc.)

# Can R Handle Large Data?

Yes! It can!

- Piping and lazy loading to avoid committing data to memory unless absolutely neccessary (***dplyr***)
- Packages that can create and communicate with relational databases (***sqldf, dplyr***)
- Partial/meta loading of data (***feather***)
- Packages to handle Large data (***big memory, ff*** etc.)
- Extensive support for parallel programming (***parallel, snow, foreach*** etc.) - allowing partitioned computation

# Can R Handle Large Data?

Yes! It can!

- Piping and lazy loading to avoid committing data to memory unless absolutely neccessary (***dplyr***)
- Packages that can create and communicate with relational databases (***sqldf, dplyr***)
- Partial/meta loading of data (***feather***)
- Packages to handle Large data (***big memory, ff*** etc.)
- Extensive support for parallel programming (***parallel, snow, foreach*** etc.) - allowing partitioned computation
- Hadoop bindings allowing management of Big data

# Statistical Analysis?

# Statistical Analysis?

- R is a statistical programming language
- Built-in functions to perform statistical computing: Modelling (regression, classification etc.) and Analysis (central tendencies, correlation, ANOVA etc.)

# Statistical Analysis?

- R is a statistical programming language
- Built-in functions to perform statistical computing: Modelling (regression, classification etc.) and Analysis (central tendencies, correlation, ANOVA etc.)

Specifically for PV:

# Statistical Analysis?

- R is a statistical programming language
- Built-in functions to perform statistical computing: Modelling (regression, classification etc.) and Analysis (central tendencies, correlation, ANOVA etc.)

Specifically for PV:

## The Pharmacovigilance Signal Detection (***PhViD***) package

# Statistical Analysis?

- R is a statistical programming language
- Built-in functions to perform statistical computing: Modelling (regression, classification etc.) and Analysis (central tendencies, correlation, ANOVA etc.)

Specifically for PV:

## The Pharmacovigilance Signal Detection (***PhViD***) package

- Functions for statistical algorithms for signal detection

# Statistical Analysis?

- R is a statistical programming language
- Built-in functions to perform statistical computing: Modelling (regression, classification etc.) and Analysis (central tendencies, correlation, ANOVA etc.)

Specifically for PV:

## The Pharmacovigilance Signal Detection (***PhViD***) package

- Functions for statistical algorithms for signal detection
- Take Drug-PT frequency table as input

# Statistical Analysis?

- R is a statistical programming language
- Built-in functions to perform statistical computing: Modelling (regression, classification etc.) and Analysis (central tendencies, correlation, ANOVA etc.)

Specifically for PV:

## The Pharmacovigilance Signal Detection (***PhViD***) package

- Functions for statistical algorithms for signal detection
- Take Drug-PT frequency table as input
- Available methods: PRR, ROR, RFET, GPS and BCPNN

# Visualization?

# Visualization?

R is well known in the data science community for its ability to create visualizations

# Visualization?

R is well known in the data science community for its ability to create visualizations

- Highly customizable and publication-ready plot generation (***ggplot2, base graphics, lattice*** etc.)

# Visualization?

R is well known in the data science community for its ability to create visualizations

- Highly customizable and publication-ready plot generation (***ggplot2, base graphics, lattice*** etc.)
- Interactive visualizations with HTML widgets (e.g. ***ggiraph***) and integration of javascript libraries like D3, plotly, highcharts etc.

# Visualization?

R is well known in the data science community for its ability to create visualizations

- Highly customizable and publication-ready plot generation (***ggplot2, base graphics, lattice*** etc.)
- Interactive visualizations with HTML widgets (e.g. ***ggiraph***) and integration of javascript libraries like D3, plotly, highcharts etc.
- Easily created with a few lines of code

# Web Deployment

R shiny: web interface for R

# Web Deployment

R shiny: web interface for R

- Simple functions to build basic components of web pages like dashboards, sidebars etc.

# Web Deployment

R shiny: web interface for R

- Simple functions to build basic components of web pages like dashboards, sidebars etc.
- Reactivity of variables and components - built-in dynamic nature

# Web Deployment

R shiny: web interface for R

- Simple functions to build basic components of web pages like dashboards, sidebars etc.
- Reactivity of variables and components - built-in dynamic nature
- Server logic can be written in R - easy representation of statistical analysis

# Summary

# Summary

- Being a statistical programming language, R comes equipped with statistical analysis functions

# Summary

- Being a statistical programming language, R comes equipped with statistical analysis functions
- R can also provide effective visualizations with added interactivity which can make data analysis more effective

# Summary

- Being a statistical programming language, R comes equipped with statistical analysis functions
- R can also provide effective visualizations with added interactivity which can make data analysis more effective
- R Shiny provides an excellent and easy-to-build platform that can take the visualizations and analysis to display online

Cytel

# Aim of the Application

Two parts :

# Aim of the Application

Two parts :

1. Database summary

# Aim of the Application

Two parts :

1. Database summary
   - Select data by specific period

# Aim of the Application

Two parts :

1. Database summary
   - Select data by specific period
   - Generate summary statistics (geometric diversity, report count etc.)

# Aim of the Application

Two parts :

1. Database summary
   - Select data by specific period
   - Generate summary statistics (geometric diversity, report count etc.)
   - Generate demographic visualizations (country wise and year wise report frequency etc.)

# Aim of the Application

Two parts :

1. Database summary
   - Select data by specific period
   - Generate summary statistics (geometric diversity, report count etc.)
   - Generate demographic visualizations (country wise and year wise report frequency etc.)
   - Demonstrate the ability of R to create visualizations from different sources (plotlyjs, d3js etc.)

# Aim of the Application

Two parts :

1. Database summary
   - Select data by specific period
   - Generate summary statistics (geometric diversity, report count etc.)
   - Generate demographic visualizations (country wise and year wise report frequency etc.)
   - Demonstrate the ability of R to create visualizations from different sources (plotlyjs, d3js etc.)
2. Signal Analysis

# Aim of the Application

Cytel

Two parts :

1. Database summary
   - Select data by specific period
   - Generate summary statistics (geometric diversity, report count etc.)
   - Generate demographic visualizations (country wise and year wise report frequency etc.)
   - Demonstrate the ability of R to create visualizations from different sources (plotlyjs, d3js etc.)
2. Signal Analysis
   - Filter out drug information

# Aim of the Application

Cytel

Two parts :

1. Database summary
   - Select data by specific period
   - Generate summary statistics (geometric diversity, report count etc.)
   - Generate demographic visualizations (country wise and year wise report frequency etc.)
   - Demonstrate the ability of R to create visualizations from different sources (plotlyjs, d3js etc.)
2. Signal Analysis
   - Filter out drug information
   - Perform statistical analysis on the data to mine for significant signals