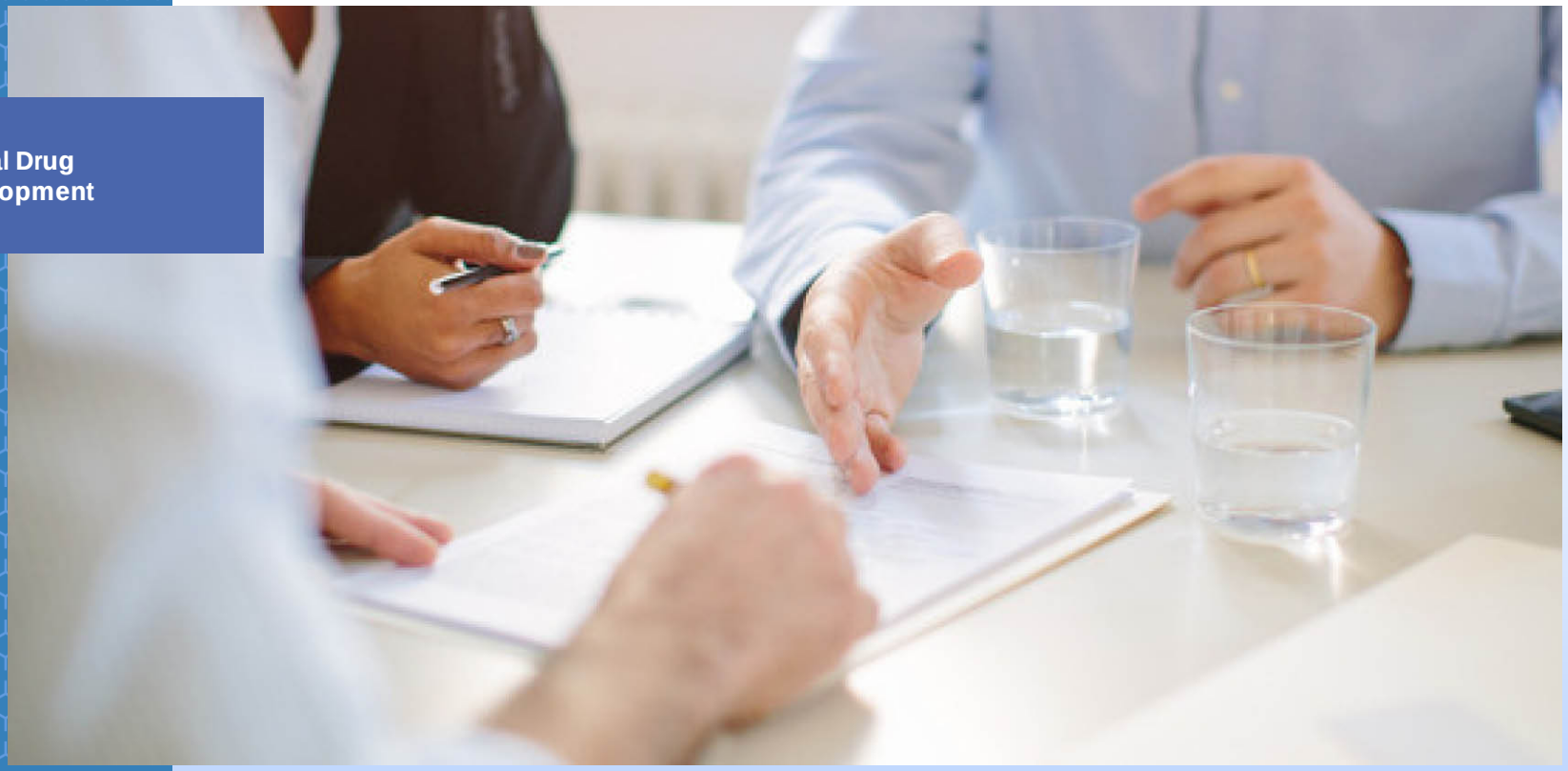
Global Drug Development



RCT to RWE
Upskilling a statistical programmer

SP Kumar / Korak Dutta
Novartis
PhUSE SDE Bengaluru May 2018

NOVARTIS

Disclaimer

- This presentation is based on publicly available information (including data relating to non-Novartis products or approaches).
- The views presented are the views of the presenter, not necessarily those of Novartis.
- These slides are intended for educational purposes only and for the personal use of the audience. These slides are not intended for wider distribution outside the intended purpose without presenter approval
- The content of this slide deck is accurate to the best of the presenter's knowledge at the time of production.

Traditional RCT

- **RCT : Randomized Clinical Trails**
- Today's established way for measuring the safety and efficacy of drugs.
- The results are obtained in
  - a highly controlled environment
  - over a small set of patient population
  - over a short period of time
  - across a handful of highly dispersed clinical trial centers

Challenges to RCT

- Payers and Health authorities asking question about sole-dependency on RCTs to establish effectiveness.
- Lack of real life evidence has driven particular increase in drug recalls, including those of blockbuster drugs such as Xigris by Eli Lilly, Avandia in 2010 (in Europe, only) by GSK, etc.

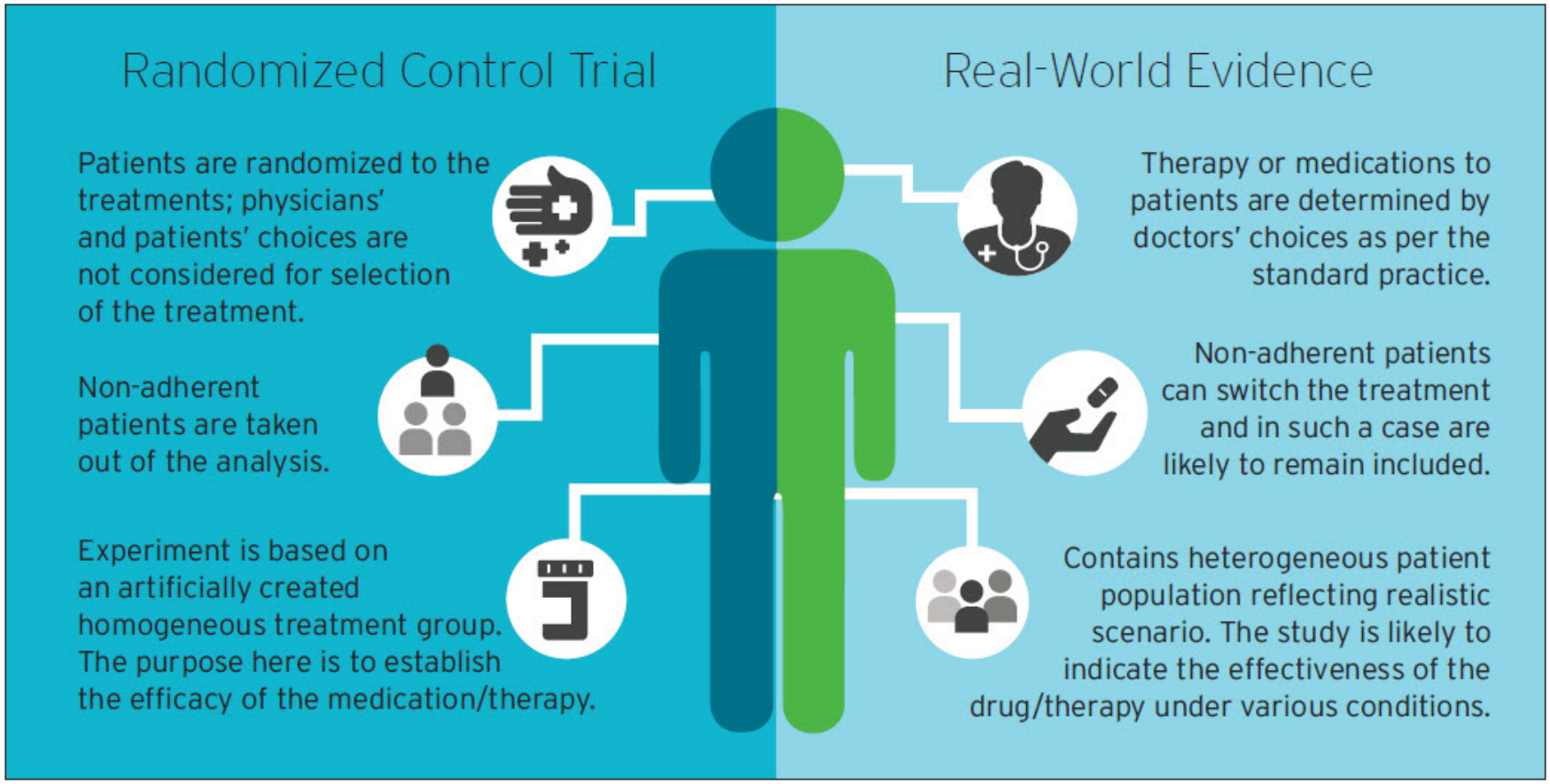
Effectiveness is the extent to which an intervention does more good than harm when provided under the usual circumstances of healthcare practice.

Evaluation of cost-efficiency of a drug in a real-world environment and measuring its impact on improving the quality of healthcare requires RCTs to be supplemented or followed up with the comparatively new standard, called real-world evidence (RWE).

RWE

- Uses observational data to generate
  - Insight
  - Foresight
- Typically, consists of four revealing data types that have overlapping and distinct characteristics
  - Predictive findings
  - patient claims data (available typically with payer) on diseases, products and patient populations.
  - patient registries (available with provider and payer)
  - electronic health records (EHR)/electronic medical records (EMR)
  - Web/social data.

RCT v RWE



NOVARTIS

RWE

Patient Claims	Patient/Medical Registries	EHR/EMR	Social Data
Data Elements			
• Patient claim information. • Patient demographic information. • Consultation details. • Hospitalization details along with cost of treatment. • Diagnoses. • Procedures/drug names with doses and days supplied.	• Patient reported data: demographic information, patient reported outcomes (PROs). • Clinician reported data: diagnosis, treatment/drug prescribed, laboratory/clinical test suggested, follow-up treatment physician rating of effectiveness. • Laboratory: diagnostic/clinical test results.	• Patient demographic information. • Patient and family disease history. • Patient physical report. • Consultation details. • Hospitalization and discharge details. • Patient operative report.	• Unstructured texts, dialogues, sentiments as expressed by the patients or their kin (in cases where the patient is managed by his or her kin).

RWE APPLICATIONS

Comparative effectiveness research
Optimal treatment algorithm
Patient adherence study
Outcome based pricing
Clinical research feasibility

NOVARTIS

Life of a typical Statistical Programmer

Provide statistical programming support

Write programs to create statistical tables, listings and graphs

Write programs to create analysis datasets

Provide functional guidance

Independently QC peer programmers' TFLs

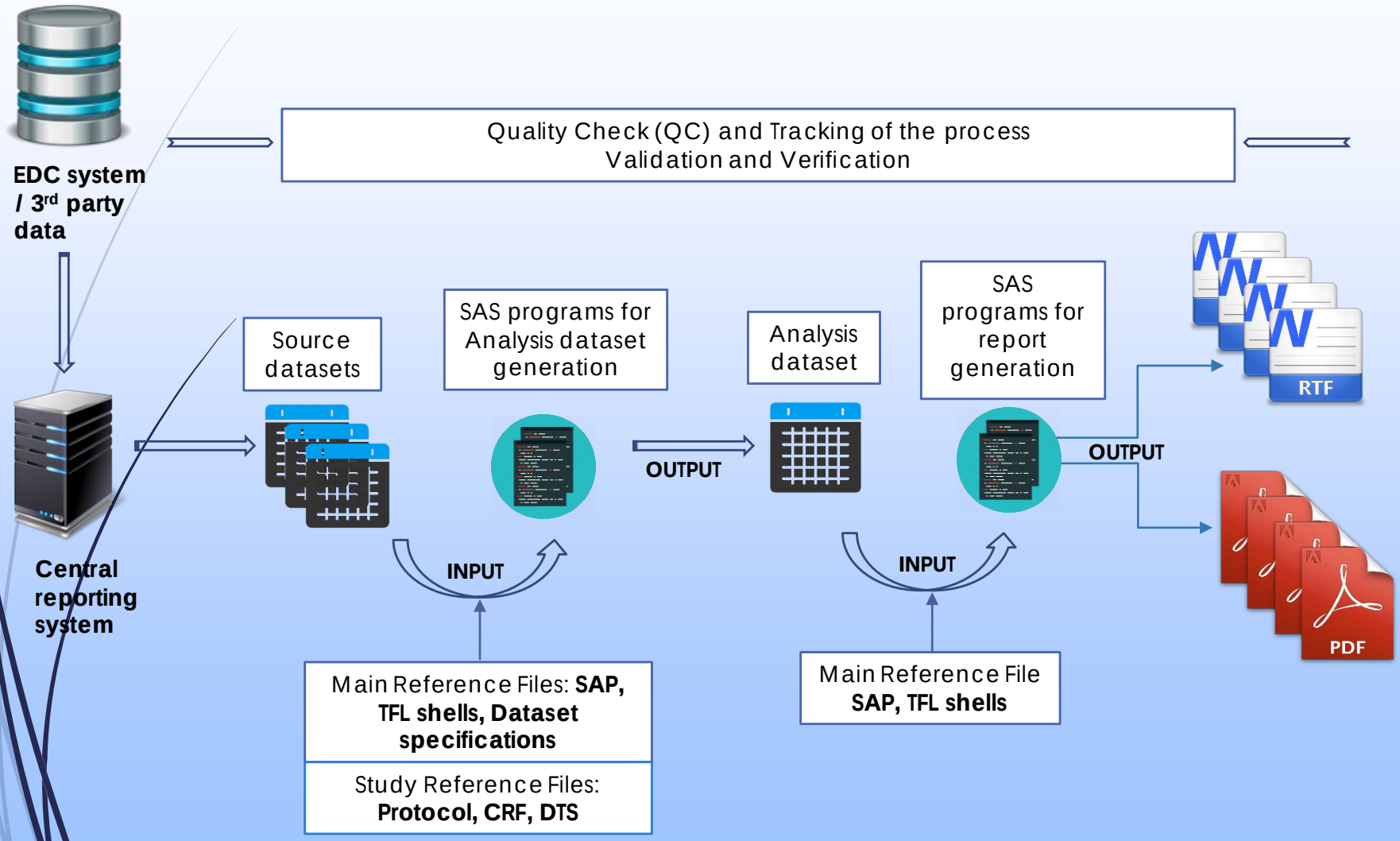
Develop operating procedures

A typical day of a statistical programmer involves developing SAS codes to create:

- Analysis datasets
- Tables
- Figures
- Listings
- Electronic submission packages
- .... And more

College degrees in Mathematics, Statistics, Applied sciences, Engineering, Computer Science, even Medicine and Pharmacy. A little programming background helps, and the role as a Statistical Programmer evolves from there.

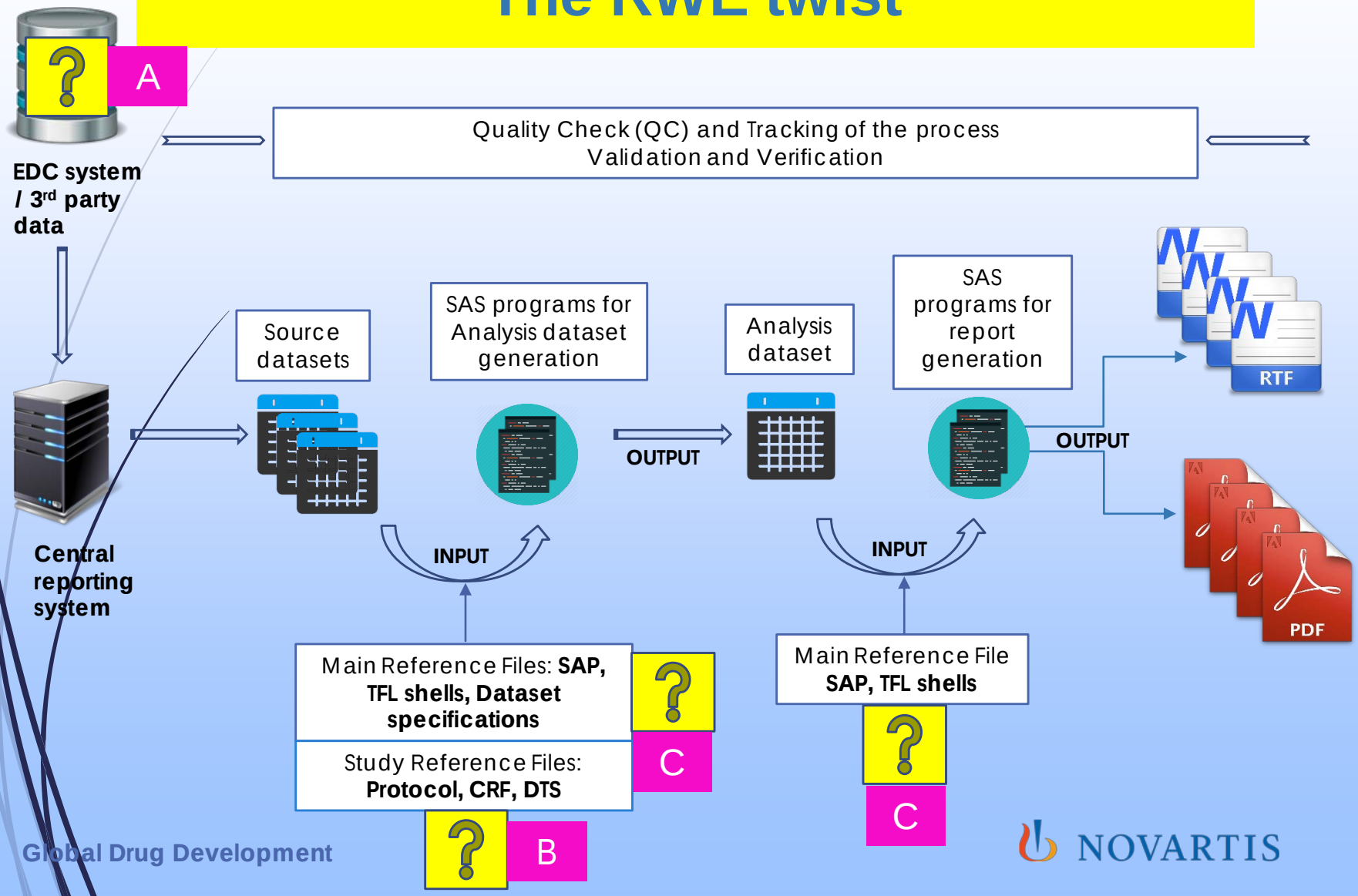
Life of a typical Statistical Programmer



NOVARTIS

Life of a typical Statistical Programmer

The RWE twist



NOVARTIS

Challenges in RWE

- A & B : Data source
- C : Analysis

Challenges in RWE

- **Data source**
- Data sources are scattered
  - social media platforms
  - claims data
  - medical records
  - published literature
- Sometimes it is organized and structured and sometimes it includes minimal text and other elements (graphics, videos, etc.).

Can be everything and anything describing in bits and bytes the journey of the patient

Challenges in RWE

- **Analysis**
- Prevailing statistical analysis models for RCTs assume a certain distributions of the sample, and uses techniques like randomization, stratifications and exclusion criterion to create a sample that can be analyzed using a model which is influenced by parameter significance.
- In search of critical growth drivers, predicting a trend and then determining non-linearities and interactions, ML techniques are more effective.

Challenges in RWE

- **Analysis**
- Due to variety of data sources and structures, ML techniques are better suited for RWE analysis.
- ML techniques are better suited where insights are to be gained using unstructured or semi-structured data.
- These techniques are very handy in the production of robust RWE by utilization data settings with highly complex mechanisms.

Some challenges for us...

- Handling and analyzing unstructured data
- Optimizing programming solutions
- Getting a handle on Machine Learning, and utilizing strengths of AI

Unstructured data

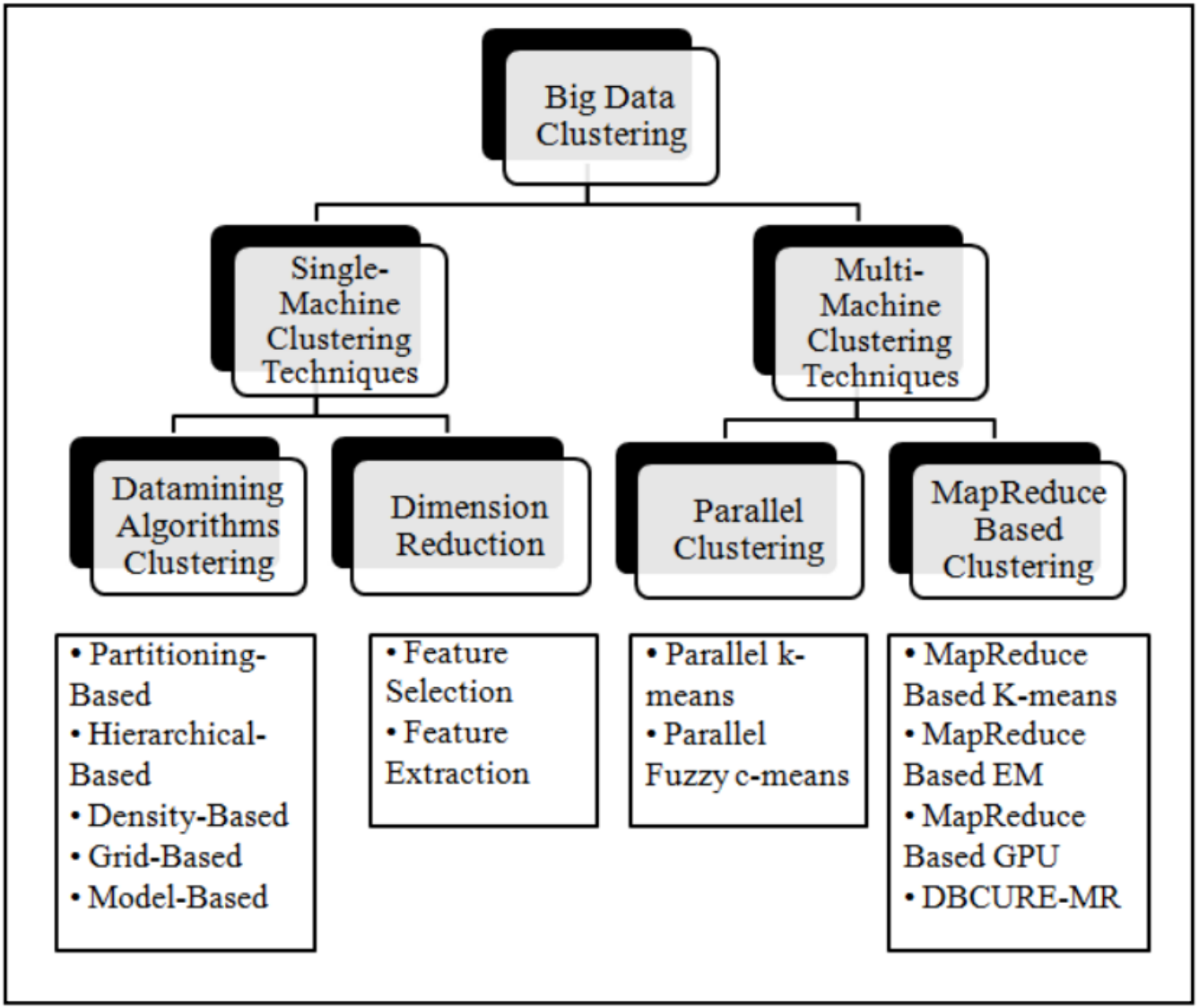
- Structured: in a database (
  Access, SQL Server, csv, Cl
- Unstructured: often hard t
  slide decks, PDFs and Wor
  interactions and blogs.
- Tools:
  - Rapid Miner (YALE)
  - WEKA
  - R
  - Orange & NTLK (Python base
  - Knime

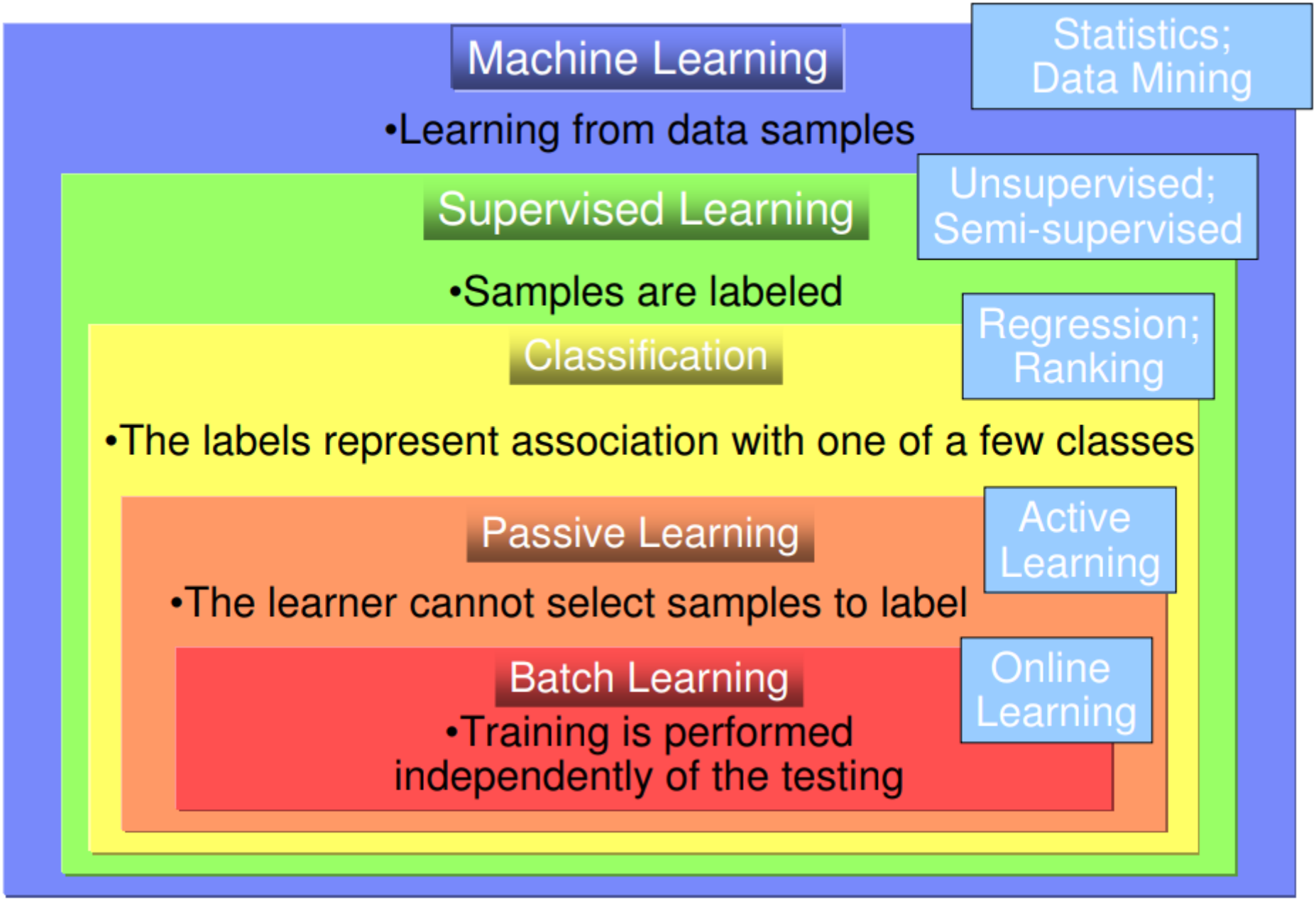
Techniques
Text analytics
Imputation
Pattern recognition
Data preprocessing
Visualization
Predictive analysis
Clustering
Association
Regression
Classification
Modelling
Time series analysis
Natural Language Processing
Data scraping

Unstructured data

- NLP: Natural Language Processing is used to preprocess and build databases from spoken or written information. Uses statistical models, 'training sets' of data, and Machine Learning methods. It is roughly divided into 4 areas of recognition:
  - Syntax
  - Semantics
  - Discourse
  - Speech

Optimizing solutions




Machine Learning
Statistics; Data Mining
•Learning from data samples
Supervised Learning
Unsupervised; Semi-supervised
•Samples are labeled
Classification
Regression; Ranking
•The labels represent association with one of a few classes
Passive Learning
Active Learning
•The learner cannot select samples to label
Batch Learning
Online Learning
•Training is performed independently of the testing

Last word

- Tools and techniques in the RWE arena are not limited to SAS
- Python, R, Perl and many other utility based languages need to be learnt to master the needs of the tasks
- Additionally, ML & AI rely heavily on statistical concepts like Bayesian models and complex theories like Graph theory. Knowledge of these is extremely necessary.

Moving from RCT to RWE requires a lot of upskilling – both in terms of programing skills as well as statistical concepts.

Dedicated time and effort is needed to upskill ourselves to this changing times.

References

https://www.invensis.net/blog/data-processing/12-data-mining-tools-techniques/

https://www.cognizant.com/whitepapers/real-world-evidence-a-better-life-journey-for-pharmas-payers-and-patients-codex1259.pdf

Zerhari, Btissam & Ait Lahcen, Ayoub & Mouline, Salma. (2015). Big Data Clustering: Algorithms and Challenges.

http://www.mlpm.eu/static/media/uploads/mlpm13_slides_rosenzvi.pdf

THANK YOU