Theme: Capabilities | Collaboration | Change on the way to Clinical Data Science

SCDM Live
India conference

2nd - 3rd December 2022
Radisson Blu Hotel, Bengaluru
IDEATANK

• Ram Mudaliar, Sr Dir, Clinical Data & Insights, AstraZeneca
• 13+ years DM experience
• India Head of Clinical Data & Insights group
• Strategic accountability for Data Management, Programming, Centralized Monitoring, Patient Safety and Data Standards, Automation in India
• Established effective DM teams in EU and North America
• Previous experience in Oncology Data Management and partnered on COVID trials with Oxford
Sowmyanarayan Srinivasan

- Sowmya has over 22 years of experience in Life Sciences working at the intersection of business and technology.
- He has worked in start ups to large consulting firms with a focus on technology transformation and innovation.
- He has set up and grown R&D technology teams and has played roles ranging from product management, business development, capability development & leading innovation.
- Currently working with Novartis as the Head of Technology Innovation & Strategic Partnerships.
Dr. Nirali Mehta

- Dr. Nirali Mehta is CEO & Founder at Pharma-Stats, India
- Trainer for Pharmaceutical Statistics across 7 countries
- Biostatistical consultant for pharmaceutical regulatory issues.
- DSMB member for vaccine trials & IDMC member for PK studies
- Consultant SME – Health intelligence, New Jersey, USA
- Regulatory consultant at Trident Biopharm Solutions UK
- Strategic Advisor for Pharmaceuticals and Medical devices at Sicrum, Pune and South Korea
Deepu Joseph

• Deepu Joseph is currently the Vice President and Country Head of Excelya India, the Indian entity of French CRO Excelya.

• With 18 years of experience in clinical research focused on Data Management and Biometrics, Deepu has worked with major CROs like Quintiles, ICON and Quanticate and has held global positions handling big and small team across the globe and different locations in India.

• Active member of SCDM India Steering committee for the last 4 years and has been contributor, speaker, and co-chair for conferences in the past
Theme: Capabilities | Collaboration | Change on the way to Clinical Data Science

"Agile Methodology in Clinical Data Sciences"

SCDM Live
India conference
2nd - 3rd December 2022
Radisson Blu Hotel, Bengaluru

Society for Clinical Data Management
Site Exposure Program (SEP)

Dr Venkateswarlu Tummalapenta
Associate Director, AstraZeneca

Nimisha Nigam
Associate Director, AstraZeneca
**Overall Process**

### Planning Phase
- **Two weeks prior to Remote Monitoring Visit**
  - SEP Champion - DM
  - SEP Champion – SM&M
  - Nominate SEP Participants
  - Site Monitor
  - Data Manager

### SEP Orientation and Participant Alignment
- **One week prior to Remote Monitoring Visit**
  - Orientation by SEP Champion - DM
  - Nominate
  - Site Monitor

### Remote Monitoring
- **Two weeks prior to Site Monitoring Visit**
  - Site Monitor
  - Data Manager
  - Data management SEP participants prepare site interaction Q&A

### Site Monitoring
- **Data Manager**
- **Site Monitor**
- **Site Staff**

### Debrief
- **Data Manager**
- **SEP Champion - DM**
- **Quarterly Meet**
- **SEP Champions Meet – DM and SM&M (SEP feed forward)**

**Engage**

- Orientation to Monitoring Visit, Site and Study details
- Guide SEP participants about what to expect during Site monitoring Visit, Do's and Don'ts
- Engage
- Engage
Business Benefits

People
- Skill enhancement of staff
- Art of story telling and listening
- Building site user experience on evolving and adoptive study designs and decentralized trials

Process
- Reduction in cancelled queries and re-queries
- People engagement through different types of site visits
- Edit check optimization (ECO)

Technology
- EHR and EDC data flow opportunity
- RAVE as a mobile application and/or Offline EDC system
- RPA BOT for eCCG

Apply Cross Functional Strengths
- CtQ (critical to Quality) data acquisition strategy
- Hand Shake of data cleaning with monitoring visit strategy
- Functional efficiencies through impactful data insights
- Engagement through patient stories
AFFILIATIONS

Dr Venkateswarlu Tummalapenta
Associate Director, Clinical Data Management CDI, Development Operations, R&D, AstraZeneca, Bengaluru, India

Nimisha Nigam
Associate Director, Clinical Data Management CDI, Development Operations, R&D, AstraZeneca, Bengaluru, India
“Integration/Amalgamation of Third-Party Data”
Industry Context

Full potential of EDC has not been realized and that most implementations ended-up converting existing inefficient paper processes inside an electronic tool.

Overall changes to downstream processes and systems were not transformational enough to take advantage of faster data availability from EDC.

Additionally, despite the faster availability of data from the sites compared to paper-based processes, the cycle time from last patient last visit to database lock has been hardly reduced over the last decade. Decreasing from about nine (9) weeks in 2008 to seven (7) weeks in 2018.

The volume of data collected outside EDC has already surpassed significantly the volume of data collected on eCRF and it is growing at much faster pace every year, driven by the cries for patient centricity leading to the rapid adoption of eCOA, wearables, sensors, TPV data and other eSource solutions.

Additionally, the increasing cost of Clinical Development and the need for greater predictability of outcome requires the use of tool which helps in real time data validation, tracking and reporting of data sources specially for high-volume data collected outside EDC in order to making a positive and meaningful impact on Clinical Development.

Source: SCDM reflection Paper
Central Data Review Tool

Queries can be pushed to EDC Real time

Extract, Transform and Load Clinical data from all sources
Generate discrepancies through programmed edit check outside EDC
Generates meaningful Operational Analytics
Outcome and Learning

Effective Data Review
- Increase data review efficiencies
- Enables Cross functional review
- Streamlines data monitoring

Targeted Data Review
- Enables in-scope data cleaning
- Projected metrics for upcoming milestones

Real Time Data Cleaning
- Generates data discrepancies throughout the lifecycle
- Reduce wait time
- Integrated Query management through EDC

Analytics
- More meaningful metrics generation and visualization to understand data trends

Automation
- Semiautomated review involves less manual intervention
- Can be enhanced with AI ML capabilities
- Realtime CPT (Clean patient Tracker)

Cost Saving
- Operational efficiencies
- Effective Resource utilization
- Reduced LSLV to DBL TAT

CDR Tool can be 21CFR Part11 and ICH-E6-R2 compliant
“Illuminating Clinical Trial Data with Actionable Analytics”
### Situation

- Searching real-time clinical data for patterns or trends for testing clinical hypothesis is currently a manual process, requiring extensive manual intervention and prone to bias

- Early (semi or fully) automated detection of important patterns or trends in study data may help clinical teams across a broad variety of TAs to adapt study conduct and/or study oversight,

### Our Approach

- A deep-learning, based hypothesis generation and prioritization module, which will leverage a range of unsupervised, deep-learning methods to identify hypothesis of potential interest, followed by synthesis and human validation

### Solution Benefits

- Enable timely interventions during trial conduct - refinements to protocol document, investigator guideline updates

- Support to review large and diverse volumes of data and efficiently identify data patterns at scale
Smart Clinical Signal Detector – Solution outline for proposed proof of concept (1/2)

Data inputs

- Data Ingestion (Historical Trial Data + Protocol + External Data Sources)
- Configuration and Data Engineering

Hypothesis Generation and Prioritization

- Configuration will include specification of input/output variables, join specifications etc.

Synthesis and Feedback

- Synthesis/Summarization (for Prioritized Hypothesis)
- Human in the loop / expert review

Validation by Clinical SMEs

Feedback to Hypothesis Generation and Prioritization Module

Insights for Future Studies
Smart Clinical Signal Detector – Solution outline for proposed proof of concept (2/2)

<table>
<thead>
<tr>
<th>Data inputs</th>
<th>Hypothesis Generation and Prioritization</th>
<th>Synthesis and Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Ingestion</strong></td>
<td><strong>Generation of Candidate Insights</strong></td>
<td><strong>Synthesis</strong></td>
</tr>
<tr>
<td>Primary Data Inputs</td>
<td>• Leveraging established and exploratory algorithms, such as</td>
<td>• Generation of summary insights based on subgroups and variables of interest</td>
</tr>
<tr>
<td>• Treatment blinded clinical trial data – SDTM, ePRO, eCOA</td>
<td>– Deep learning based phenotyping methods</td>
<td></td>
</tr>
<tr>
<td>• Protocol document</td>
<td>– Generalized mixed effects models</td>
<td></td>
</tr>
<tr>
<td>• Known safety concerns</td>
<td>– Evolutionary Search Algorithms</td>
<td></td>
</tr>
<tr>
<td><strong>Reference Data Inputs</strong></td>
<td>– Recursive tree-based partitioning</td>
<td></td>
</tr>
<tr>
<td>• Historic column mappings and domain ontologies</td>
<td>– Exploratory algorithms including Deep Reinforcement Learning, Deep Learning, Frequent pattern mining etc.(^1)(^2)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Configuration &amp; Data Engineering</th>
<th>Prioritization of candidate insights</th>
<th>Validation and Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Configuration (Human expert inputs)</td>
<td>• Lift, Statistical significance tests</td>
<td><strong>Synthesis/Summarization</strong></td>
</tr>
<tr>
<td>• Primary, secondary end-points identification</td>
<td>• Sub-group effect size estimation and comparison vs. global threshold</td>
<td><strong>Synthesis</strong></td>
</tr>
<tr>
<td>• Column mappings using human expert inputs/ ontologies</td>
<td>• Reward functions defined on the basis of interestingness, diversity and coherence</td>
<td>Summary Insights available for review</td>
</tr>
<tr>
<td><strong>Data Engineering</strong></td>
<td><strong>Feedback to Hypothesis Generation and Prioritization Module</strong></td>
<td><strong>Validation by Clinical SMEs</strong></td>
</tr>
<tr>
<td>• Generation of multiple databases, as needed, combining identifiers, patient demographics, eCOA responses, safety and efficacy indicators, at a patient-time or patient level as needed</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Case Study – an ensemble-based approach to identify historical trends or patterns in various studies and detecting safety and efficacy signals

Situation

- Client has near real-time clinical data like CRF, PRO etc. and currently rely on manual data review process which is time-consuming and laborious; wanted an AI/Machine Learning based technology platform with the ability to surface important and clinically meaningful trends or patterns during study conduct, such as
  - Cluster patients into treatment/placebo cohorts based on treatment response
  - Early prediction of treatment response
  - Identify safety signals

Outcomes

- Was able to identify patients with high/low response early in the trial (utilizing data available in the first snapshot), along with drivers; subsequent data improved the accuracy
- Uncovered clinically meaningful trends or patterns during study conduct, such as (but not limited to) -
  - High response in patients with a high dose of NSAIDs
  - High Probability of TEAE such as Arthralgia, Osteoarthritis, Back Pain, associated with ongoing medications such as ASA, Supplements, Chondroitin/Glucosamine, Levothyroxine, etc.
  - Correlation between liver adverse event, and treatment, possibly due to non-linear combination of concomitant medication and the treatment in question

Key Learnings

- Focus on explainability key as clinical SMEs are key consumers – use of effective visualizations key
- Ensemble of methods needed to capture diverse signals of interest across TAs
Approach #1 - Computational phenotyping for identifying insights in complex, multi-dimensional temporal data

Vector Representation
Numerical representations of individual patient/visit level data are automatically generated using embedding techniques to help render a digital portrait enabling unsupervised subtyping via graph modularity or other clustering methods.

Pattern Detection
Clusters in the representations are identified, which determine why certain groups of patients lie in a tight vector space distinct from others and further what signals are of importance within each of these clusters (potential trends/patterns).

Post-profiling & Validation
Descriptive summaries to aid in hypothesis prioritization.
- Cohort characteristics to understand patterns and trends from blinded trial data.

- CNNs to model temporal aspects coupled with autoencoders to enable unsupervised architectures (ConvAE).
- Tanh-LSTM based encoder-decoder architectures.
- Stacked denoising autoencoders (DeepPatient) architecture.
- Graph-based methods which leverage hidden structures in the data.
Approach #2 - Proposed exploratory method for an Automated Insight Engine, utilizing optimized frequent pattern mining

This proposed exploratory approach aims to extract top insights from any transactional dataset, using optimized frequent pattern mining methods – the approach requires configuration of data taxonomy, necessary operations and scoring functions to automatically traverse through potential insights and highlight top statistically significant insights.