Table 22: Big Data Applications in Biopharmaceutical Process and Product Development

Facilitator: Kshitij Khatri, *GlaxoSmithKline, Collegeville, PA, USA*

Scope:

The world today revolves around information – the past decade has seen a revolution in advancement of technologies enabling the generation, management, analysis, and dissemination of data. The biopharmaceutical industry can gain significantly from proper utilization of chemometric data associated with process and product development. This roundtable will discuss the potential of and challenges around use of "big data" strategies toward decision making and acceleration in biopharmaceutical CMC.

Bioprocess development requires constant monitoring of process parameters that are routinely correlated with product quality attributes. These correlations over the course of multiple process and product development campaigns can be used to identify controls for manufacturing platforms. Further, knowledge gained from such campaigns, and implementation of data science around chemometrics enables meaningful interactions with the regulators on quality-related topics.

One of the greatest challenges in exploiting big data is around harmonization of data structures – both within biopharmaceutical companies, and across the industry. How do we standardize big data in the biopharmaceutical CMC space, and apply informatics and machine learning strategies for predictive modeling toward CMC acceleration? What is the regulators' perspective on use of big data strategies and use of prior knowledge from platform processes toward drug approval?

Questions for Discussion:

1. What are the desirable features in data management platforms? Are there preferred vendor platforms for data management that innovators utilize? Discuss challenges with platforms and gather suggestions toward improvement.
2. What would it take to establish standards in data structuring and management across biopharmaceutical companies – Industry consortia and regulatory buy-in?
3. What types of predictive modeling and/or decision making (applications) do innovators typically utilize big-data strategies for today? What other applications do we envision?
4. What sorts of data or prior-knowledge driven inferential outcomes do the regulators typically encounter? What is their comfort level and expectation with these approaches being used toward demonstrating quality (safety and efficacy)? Are there regulatory guidelines around this? Share examples and discuss situations.

Discussion Notes:

1. What are the desirable features in data management platforms? Are there preferred vendor platforms for data management that innovators utilize? Discuss challenges with platforms and gather suggestions toward improvement.
    a. Compiling analytical data for characterization can be very time consuming. Having tools that can be used to manage those large datasets can have a huge advantage.
    b. There is a wealth of data generated for each molecule and it's possible to streamline process dev if we have adequate data management tools.
    c. Ex: HCP proteomics to determine how each unit op impacts HCPs
    d. Challenges:
    e. Appropriate/adequate meta-data built within the datasets.
    f. Cleaning and organizing the data.
    g. People don't like change. Data hygiene is lacking across teams.
    h. Who drives the change? Scientists or leadership?

2. What would it take to establish standards in data structuring and management across biopharmaceutical companies a" Industry consortia and regulatory buy-in?
    a. No one really had an answer for this.

3. What types of predictive modeling and/or decision making (applications) do innovators typically utilize big-data strategies for today? What other applications do we envision?
    a. Some participants have built their own systems and others use some off-the-shelf software. Doesn't seem like
    b. Leveraging prior knowledge to develop risk assessments. How are we keeping track of prior knowledge? Is there a place for people to download their own knowledge into a local space where it can be accessed across the company or industry? How do you encourage people to share their knowledge?
    c. No one seems to be using prior knowledge to replace actual development work but it can simply the development work.

4. What sorts of data or prior-knowledge driven inferential outcomes do the regulators typically encounter? What is their comfort level and expectation with these approaches being used toward demonstrating quality (safety and efficacy)? Are there regulatory guidelines around this? Share examples and discuss situations.
    a. iRisk – risk management software. Can be customized depending on what companies want to use it for. You can build workflows to store outcomes and then reference it within the system. It will help to determine risk scores.
    b. What data language/algorithms do people use? One person uses machine learning to build predictive models but it's still in early stages.
    c. Are there parallels that can be drawn to see how other departments/sectors (i.e., sales) manage data?