BioCelerate SEND Harmonization Initiative – Public Workshop | *Minutes*

DATE/TIME: **March 13, 11:00 – 01:00 PM EDT**

LOCATION: **Hilton Baltimore, Society of Toxicology 2019 Annual Meeting (Ancillary Event)**

ATTENDEES:

Representatives from the BioCelerate SEND Harmonization Team:

- Thomas Bjerregaard, Toxicology Specialist, CDISC SEND Expert (Novo Nordisk) & Initiative Team Member (BioCelerate)
- Mark Carfagna, Senior Research Advisor, Toxicology (Eli Lilly) & Initiative Team Member (BioCelerate)
- Tamio Fukushima, Director (Shionogi) & Initiative Team Member (BioCelerate)
- William Houser, Principal Scientist (Bristol-Myers Squibb), SEND Team Leader (CDISC) & SEND Harmonization Lead (BioCelerate)
- Kelsey Jakee, Director of Portfolio (BioCelerate) & Principal Consultant, Life Sciences (PA Consulting)
- Raja Mangipudy, Executive Director, Toxicology (Bristol-Myers Squibb) & Data Sharing Initiative Lead (BioCelerate)
- Todd Page, Director, Toxicology (Eli Lilly) & Nonclinical Study Optimization Initiative Lead (BioCelerate)
- Cheryl Sloan, Application Manager & Research Scientist (Bristol-Myers Squibb) & Initiative Team Member (BioCelerate)

Other individuals present at the meeting represented biopharmaceutical companies, FDA, software/platform providers, contract research organizations, and other industry service providers.

AGENDA TOPICS:

Topic	Timing	Leader
Introduction & Overview	11:00 – 11:25 am	Todd Page
Lunch & Networking	11:25 – 11:45 am	Individual Tables
Small Group (Table) Discussion	11:45 – 12:15 pm	Individual Tables
Full Group (Room) Readout	12:15 – 12:45 pm	Raja Mangipudy
Closing & Next Steps	12:45 – 1:00 pm	Raja Mangipudy

INTRODUCTION & OVERVIEW:

Todd Page provided an introductory presentation on BioCelerate, its portfolio, and the purpose of the SEND Harmonization initiative. BioCelerate is a subsidiary of TransCelerate BioPharma, Inc., a not-for-profit entity created to foster collaboration. BioCelerate's mission is to identify, develop and deliver industry initiatives aimed at making nonclinical R&D more effective.

The Toxicology & Background Control Data Sharing initiative has delivered a platform (DataCelerate) to enable the sharing, search, and visualization of de-identified toxicology and background control datasets submitted by participating companies. The DataCelerate platform will also contain de-identified clinical Placebo & Standard of Care datasets contributed by TransCelerate member companies beginning in Q2 2019. The Nonclinical Study Optimization initiative is supporting two projects, Common Templates for Nonclinical Studies and SEND Implementation for Study Analysis.

The SEND project is primarily aimed at facilitating greater harmonization of the CDISC SEND standard to enable various cross-study analysis use cases. In a recent webinar poll of various stakeholders spanning biopharmaceutical industry, CROs, technology vendors, health authorities, other service providers, and industry consortia, respondents were asked which of the following five use cases was deemed most valuable using harmonized data sets (results indicated):

- Understanding the toxicity profile of all studies per compound (58%)
- Understanding off-target toxicity or trends per class/MOAs (31%)
- Understanding frequency of rare/incidental findings in control data (11%)
- Understanding the effects of vehicles (0%)
- Enabling application of SEND data within QSAR (0%)

It was discussed that due to the phrasing of the question and the requirement to select just one use case, three use cases rose to the top of the list as being considered a first 'top priority'; however, the remaining use cases are valuable and were likely not selected due to being a second or third top priority. It should be noted that the QSAR use case is considered a more difficult ambition that is unlikely to be realized in the near term.

A second webinar poll also identified relative weight on the significance of various drivers causing variability across SEND datasets:

- Standard operating procedures that vary across companies (7%)
- Protocols/study plans of varied scope and nomenclature (37%)
- LIMS systems with varying software packages and set-up (19%)
- Flexibility in the SEND standard itself (37%)

BioCelerate is proposing a collaborative, four-stage approach to develop future solutions for the harmonization of SEND that enables various high-priority cross-study analysis use cases. Involvement is requested from industry, CROs, CDISC, health authorities, and technology vendors, to understand drivers of variability, prioritize focus areas for alignment, and identify/develop potential solutions.

LUNCH & NETWORKING:

SMALL GROUP/TABLE DISCUSSIONS

Four tables of 7-9 attendees each gathered to engage in small group discussions on three priority questions related to the future harmonization of SEND:

- What additional cross-study analysis use cases for SEND harmonization have BioCelerate not yet considered that are valuable across stakeholder groups?
- What are the expected benefits and challenges of SEND harmonization? What are the biggest issues or drivers of variability?
- How can we approach next steps in a collaborative way? What is the right path forward?

FULL GROUP/ROOM READOUT

Use Cases for SEND Harmonization / Cross Study Analysis

It was acknowledged that SDTM and SEND datasets are substantially similar and could be linked together to drive translational value across preclinical and clinical data. Analysis across data sets could enable new insight into findings of clinical importance that are discovered earlier in preclinical studies.

Additional uses cases discussed by the group include:

- Looking for correlations between findings in clinical trials and toxicology findings in animal studies
- Identifying across a variety of compounds a specific finding observed in one (reference) compound
- Integrating -omics data with toxicology data
- Supporting in silico mining / network analysis (interpret analyses by mapping to raw SEND data)
  - Can start assessing validity of study data based on larger network analysis
- Improving operational efficiency by analyzing historical (meta)data to identify opportunities for continuous improvement

It was noted that the original use case of understanding toxicity profile of a compound across multiple studies is very broad and can encompass:

- Visualizing data in different ways – e.g. looking across all data for a single compound through time (longitudinal analysis)
- Understanding how toxicity profiles play out at a basic scientific level – e.g. what does liver toxicity generally look like and what are the common patterns (linking clinical pathology to histopathology)
- Systems-level toxicology

It was also noted that generally, study directors would like to be able to look across many studies in a more flexible manner – to better understand historical control data, institutional control data, to visualize multiple studies across compounds, etc. Those working on R&D pharmacology studies often do not have access to SEND information from GLP studies.

Benefits / Challenges (including drivers of variability)

The table groups agreed there is value in harmonizing SEND data sets to enable broader, population-level information and cross-study analyses. In hindsight, the flexibility inherent in the SEND standard could possibly have been narrowed further to account for such use cases. It is not always obvious on a study-by-study level that variability is a problem, but it becomes more obvious as an organization compiles more datasets from different CROs, etc. and even from disparate sources internal to the same organization. Additionally, data mining of such datasets requires a level of harmonization that does not exist today.

Various challenges were discussed and presented as either a barrier to harmonization or a potential target for harmonization:

- There is often a lack of transparency, even within individual companies, about how SEND datasets are compiled or interpreted
- There is a lot of subjectivity in the interpretation of the SEND Implementation Guide
- There is a concern around putting too many limitations on the SEND standard – it was designed with flexibility in mind
- On a practical level, organizations are trying to deliver SEND packages as quickly as possible and in many cases are struggling to keep up with new demand; there is less time to think long term about how we want to analyze data in the future
- Significant manual intervention is currently required to make datasets more harmonized, indicating this is not a trivial issue (e.g. grouping different sets together and identifying which ones are toxicokinetic or recovery groups)
- Specific terminologies have become quite complicated (pointing to a need for more controlled data) and some controlled terminology lists are extensible (therefore terminology may not be consistent across studies)
- The existence of wide variability indicates multiple drivers may be at play, some of which may be "hard coded" into existing platforms (e.g. how LIMS systems develop the data)
- For many studies there are multiple valid ways the SEND trial design could be populated
- How to capture what the analysis really means, today versus years from now
    - Data drift and trending over time: commonly the historical control data within 5 years of a specific study's execution is considered relevant for the evaluation of that study
    - Legacy data may have little use in the future

Additional challenges with specific SEND domains and variables were also discussed:

- TPT timing variables: some measurements are expected to change relatively quickly following a dose. The free-text label for the time of the measurement following the dose is represented in the --TPT variables in SEND.

- Study ID: each entity involved in the study has a unique Study ID. Each sponsor has varying requirements of what ID to use
- CL / clinical observations: SEND imposes little structure to the clinical observations domain. As a result, there is wide variability across implementations. The scientific community has begun to improve upon this, particularly the community of SEND users in Japan.
- Limit of Quantitation (LOQs): The FDA Technical Conformance Guide (TCG) now requests BLQ to be supplied in PCSTRESC for LLOQ and the lower limit of quantitation to be supplied in PCLLOQ. Guidance is not imposed for other domains, upper limits of quantitation. Also, guidance is lacking to know if PCLLOQ is the method's limit or the sample's limit.
- Reference ranges: it would be preferable to have reference ranges for certain parameters (e.g. for certain species)

Thoughts on Next Steps

Value Proposition / Vision

- There would be value in building an ontology to link MedDRA with controlled terminology for histopathology
- It is important to understand ways to build substantial volume of studies in a central location for analysis
  - Technology is not challenging; getting people to contribute is
  - Cleaning and harmonizing the data is not trivial – this is a value-added function
  - Demonstrate real world value cases to build support for the idea of data sharing - people aren't believers until they see a real-world story of value
- End-to-end data linkages across the lifecycle of a study are key
  - Ability to link the protocol to the study dataset and the report, with harmonized and traceable metadata across the lifecycle, requires that we meaningfully consider what metadata are most needed
  - Any warehouse without enough metadata to describe the studies fully will not be truly valuable in the long run
  - Ultimately, SEND datasets should accurately reflect the study report

Stakeholders

- Smaller companies may not see as much of a benefit in cross-study analysis because they likely aren't using a single database for their studies and/or don't have a critical mass of studies for analysis – this may be led by people and organizations with more to gain from sophisticated data warehousing and analysis
- CROs need to involve both their business stakeholders and their technical stakeholders in this effort
- There is added value in BioCelerate offering a mechanism/platform that helps to automatically curate the data submitted from multiple companies, thus supplying a single location that houses harmonized versions of SEND data sets
- Bring other expertise from other data warehouses – where they've succeeded, failed, etc.

- How did others get over barriers of contributing studies from for-profits?
- Past efforts that have succeeded may have done so by involving academics

- The FDA may be able to provide a leadership role on multiple fronts
  - The FDA could communicate a harmonized approach in a way that is better enforced – helping industry move to an environment where the protocol, dataset and submission are all part of the same pathway
  - If sponsors approached the FDA with a willingness to share their blinded data for broad, cross-study analysis across industry, more innovative approaches could be entertained
    - Under current regulations, FDA does not own the data – they just receive it – it is the sponsor's proprietary data and the FDA cannot use it for other purposes
    - Similar approach applies to CROs who handle sponsor data – walls separate each client's datasets

(end)