Beyond SEND:

leveraging non-clinical data to drive translational research forward

The Pistoia Alliance (www.pistoiaalliance.org), a not-for-profit collaboration of life science companies, vendors, publishers and academic groups that work together to lower barriers to innovation in R&D, brought together a cross-functional life science industry group to review common challenges and opportunities in non-clinical drug development. The group focused their discussion on the challenges and opportunities for leveraging non-clinical datasets in support of translational research. This paper provides a summary of their opinions and conclusions.

By Katherine Briggs, Bob Friedman, Kristen Ferrara, Mark Pinches, Will Drewe, Cheryl Riel, Melody Thompson, Antoinette Hayes, Jason Gratt, Shylah Wyllie and MaryBeth Walsh

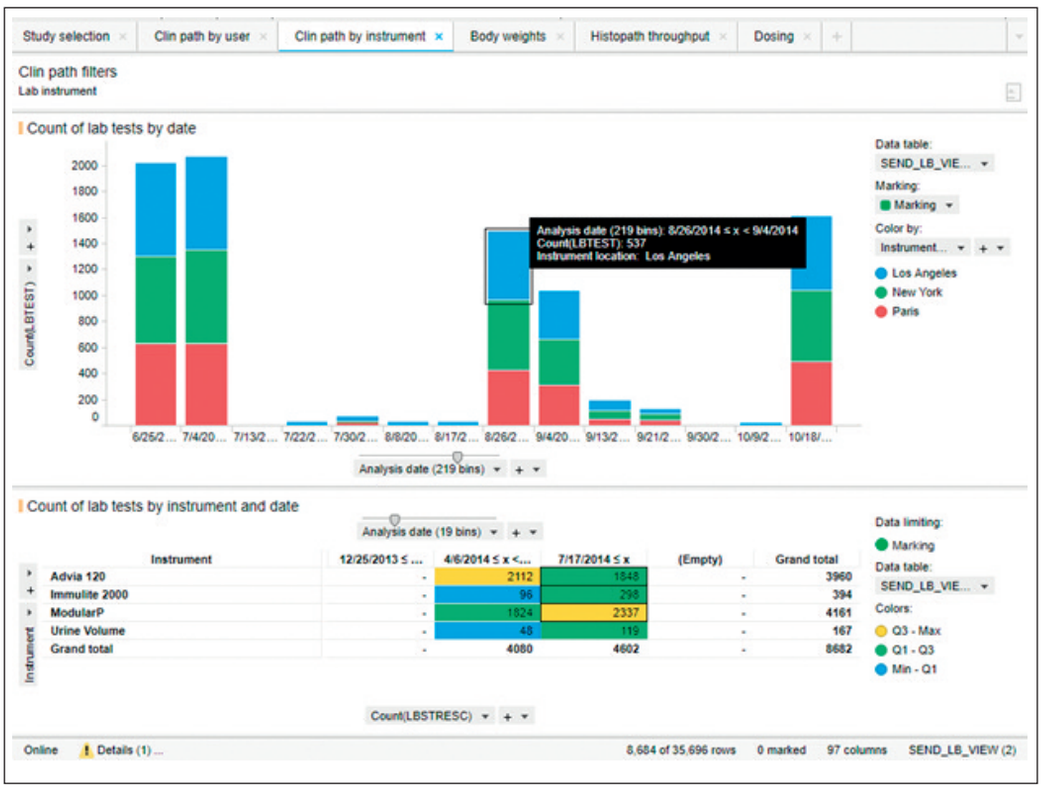
The Standard for Exchange of Nonclinical Data (SEND) is one of the required standards for data submission to the FDA and specifies a way to collect and present non-clinical data in a consistent format. For the industry, standardisation on the SEND format offers an opportunity to move beyond submission readiness to extract significant scientific and operational insights from the data. Increasingly, organisations are recognising the potential of leveraging this high-value information for improving research operations both internally and with partners via data mining, visualisations and AI/advanced analytics. This paper will discuss some of the potential use cases that SEND datasets can support, both to advance translational research across multiple research sites and also to drive efficiencies with research partners including contract research organisations (CROs). In addition, the authors will explore applications of consolidated, precompetitive data sharing to support cross-industry safety and toxicology applications.

SEND as a regulatory data submission format

The Standard for Exchange of Nonclinical Data Implementation Guide (SENDIG)[1] was developed by the Clinical Data Interchange Standards Consortium (CDISC). SENDIG is intended to provide guidance for the organisation, structure and format of non-clinical tabulation datasets intended for regulatory submission and for exchange between organisations. Pharmaceutical companies and CROs are now required to submit SEND datasets for certain types of studies when filing Investigational New Drugs (INDs) and New Drug Applications (NDAs)[2]. SENDIG sets the standard for non-clinical dataset files containing electronic records of protocol design, animal demographics, animal exposure and animal observation data and detailing their content and terminology. SENDIG Version 3.0 applies to single-dose and repeat-dose general toxicology and carcinogenicity studies. SENDIG Version 3.1 adds standards for the electronic data records for certain types of safety pharmacology studies and SENDIG Developmental and Reproductive Toxicology (DART) Version 1.1 adds standards for embryo-fetal developmental toxicity studies.

By creating a standard that is now required for submission of data, CDISC SEND is streamlining the communication between CROs, sponsors and FDA during the process of study conduct as well as at time of submission. CROs are increasingly using these formats to send interim datasets to sponsors.

Figure 1
Operational metrics in a clinical pathology laboratory[4]



Sponsors can review preliminary data from their studies and perform early analysis of the results. More and more, sponsors are looking to CROs to provide SEND-compliant datasets to simplify the submission process. For CROs, implementing the systems and processes to produce SEND-compliant datasets can provide an important differentiator in a highly competitive market.

Looking beyond the benefits of a common data model and its ability to provide the required datasets to regulatory agencies in the required format, SEND presents an important opportunity for biopharmaceutical R&D. As data complexity increases and the need for new therapies to meet unmet medical needs continues, SEND datasets, if leveraged properly, have the potential to provide unique operational and R&D insights. Such insights can increase efficiencies, reduce failure rates, and improve safety outcomes in the drug development lifecycle. Considering that one drug can cost an organisation nearly $3 billion to develop to market, the leveraging of SEND data is an important means to increasing the return on an existing investment[3].

SEND and data mining, visualisation and advanced analytics

Example 1: Operational metrics

Operational metrics are used by a laboratory to gauge the efficiency of its operations over time. An example of operational metrics tracked by a typical clinical pathology lab are illustrated in **Figure 1**. This visualisation relies upon the count of records in the lab domain (LBTEST) over time (LBDTC) grouped by instrument name and instrument location. A convention of using the LBMETHOD raw data to include a concatenation of the instrument name and instrument location makes this possible.

A stacked bar chart is then used to show the distribution of the work among different sites or laboratories where the analytical instruments exist. This can be useful to determine over time the workload taken on by each site within the laboratory network. The Y axis represents the number of tests done in that time period. The time period can be set by the visualisation tool in blocks of different sizes and a particular time period can be zoomed into in order to be specific to that time period in the analysis.

Below this is shown the same information grouped by instrument. This tells us which type of instrument is responsible for the bulk of the analysis during time periods. A trend over time provides information on which additional instruments may be needed to anticipate capacity issues. The visualisation tools allow one to set the size of the date periods, in order to look at a grosser or finer dis-

tribution of the data over time. The colouring is set to show where the maximum and minimum usages are occurring.

The visualisation tool allows filtering by other fields that are available in the SEND data. For example, filtering by study type and species to ascertain the relationship of the number of tests performed by location and instrument for each type of species and/or study type. Such operational information is most useful for CROs which are doing high volume work and need to plan their equipment resources to align with incoming studies. Biopharmaceutical companies that are performing in-house studies would also benefit from metrics on their laboratories' throughput. Companies that outsource most or all of their studies could gain insight that would help them in future CRO selection and study monitoring.

In the past, data on laboratory metrics may only have been available for individual studies. Now a data warehouse of SEND datasets opens up the possibility to gain an overall view of the data history for many studies over time. This allows one to look for trends in types of analyses and where they are being carried out. This knowledge can be used to impact the cost and timeline for future study completion.

Example 2: Data mining

In research it is useful to have the flexibility to examine the data in an open-ended fashion – searching for patterns and correlations. The ability to review any of the in-life, necropsy and histopathology observations across one or multiple studies is another benefit brought about by the standard representation of the data through SEND datasets (see **Figure 2**).

In this visualisation, clinical pathology, organ weights and micropathology data are available for filtering and exploring to search for meaningful correlations. Each section is coloured or separated by dose group.

The controlled terminology standards being developed for neoplasm and non-neoplasm allow cross-study comparison that was previously difficult due to differing terms as well as different ways to combine base terminology with modifiers.

The heat map is just one way to see the dimensions represented in this graphic. Stacked bar graphs show the data separated horizontally or vertically which may be easier for some pathologists to see patterns that need further exploring. Scatter diagrams are often easier to see clusters of information from individual observations. Sunburst plots are another way of representing the incidence data as a series of concentric arcs, where

Figure 2
Visualisation of pathology data

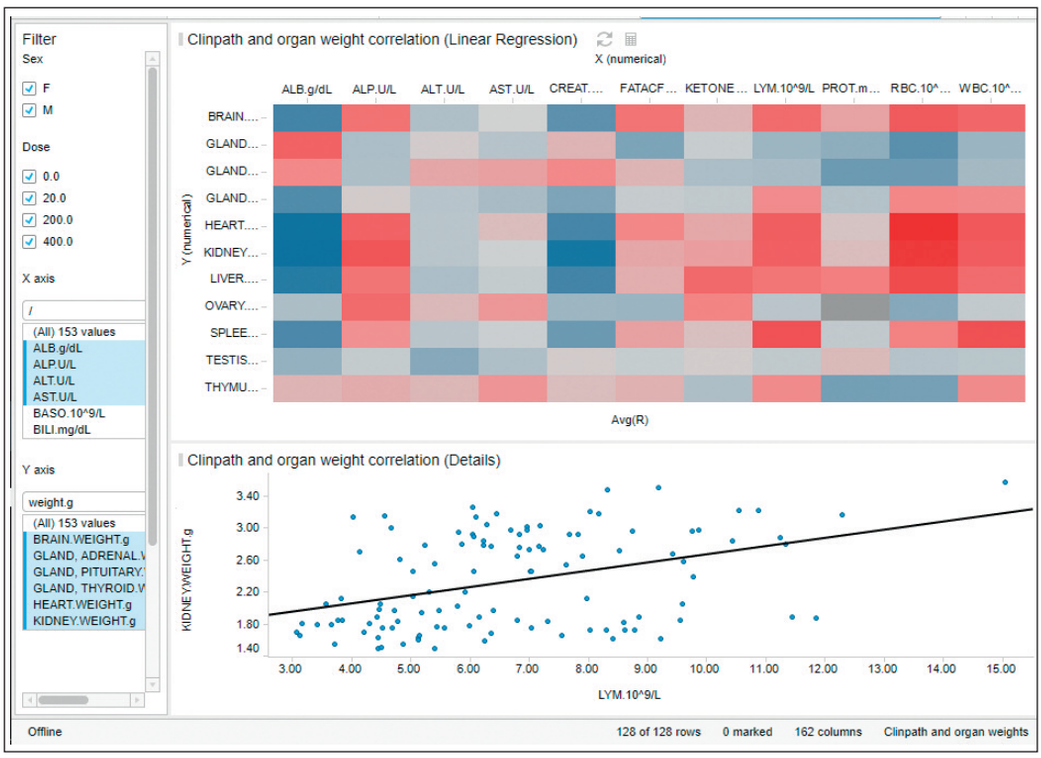


Figure 3
Visualising correlations in pathology data

one can zoom in or zoom out on the data represented from body systems, to tissues, to findings. Some may prefer to see this data represented as numbers in a table with the added benefit of conditional colouring to draw one's eye to incidences higher or lower than the norm.

Calculating correlation statistics is another ability of such tools. For example, clinical pathology data can be correlated against organ weight data to help the scientist concentrate analysis on areas of high correlation.

All of these visualisation types are being explored as pathologists are now having such data readily made available to them. The goal is to empower the pathologists to take full advantage of the standardised data across studies, compounds and animal models in order to find toxicologic effects or candidate treatments for disease targets.

Example 3: Study monitoring

Study monitoring and study analysis can also be enhanced by the ability of SEND datasets to be used with a visualisation tool. The standardised data types and terms and the ability to receive in a dataset all the measurement types on a study make possible the development of tools to view the data in ways that the study director finds most conducive to their analysis.

This graphic (**Figure 3**) shows clinical pathology, clinical observations, body weights and micropathology on the same page.

The clinical pathology is displayed as a scatter diagram with time as the X axis (the study visit day, ie the scheduled study day is used for time) and the test value as the Y axis. The clinical observation data is shown as a sparkline, with time along the X axis and incidence as the Y axis. Each sparkline is for a different clinical observation.

The body weight data is also shown as a scatter diagram, with time as the X axis, using the same timescale as the clinical pathology data. The Y axis is the body weight value. Colour is used to show the different dose groups. The micropathology data is shown as a heat map. Each section down the graph represents a different dose group. The map is categorised by tissue and then the standard finding terminology.

The charts are linked in that selections on one graph change the data displayed in the other graphs. This is done so that a selection of clinical pathology data points of interest will then filter the clinical observations, body weights and micropathology findings to those same animals.

If the lab performing the study has the ability to deliver interim SEND datasets, the sponsor can view the data as the study progresses. This can be useful in looking for test article-related effects or adverse events that might show up early in the

study if the doses are too high, allowing for an early remediation if needed during the study conduct.

Example 4: Advanced analytics and predictive modelling

A variety of opportunities to perform more advanced analytics and predictive modelling emerge if the challenges of study formatting and data capture into large databases are overcome. These opportunities range from implementing traditional statistical methods in an automated manner, to more sophisticated data-led and machine learning techniques. From a single study perspective, novel implementations could include automated methods to surface anomalous findings, generating dose response models for anomalous findings, and fitting simple pharmacokinetic and pharmacodynamic models to study data. When multiple studies are available, then emphasis can shift to either analysis of multiple compounds against any given study end point, or to a study end point specific focus, where the behaviour of endpoints can be investigated. Hierarchical dose response models, that fit end point specific data for multiple compounds in a single model and allow subsequent ordering of potency have been developed but are not yet available commercially. In addition, endpoint specific analysis that reveals, for instance, the background incidence rate of histopathology findings or putative maximum physiological responses for laboratory tests, have already been performed on large datasets of preclinical data, but not yet on data formatted in the SEND standard (see **Figure 4**).

With the emergence of these approaches, enabled by adoption of SEND as an exchange format, the field will move towards increasing automation and more data-led toxicological predictions. Since data will be stored consistently in databases, it will become easier to generate data representations for any given compound that link compound with target (primary and secondary) and with perturbed preclinical endpoints. This enables toxicology to engage with system biology approaches that offer another approach to the prediction of clinical adverse events.

Barriers to advanced analytics

The tools are available to leverage cloud resources to share standard formatted data across research sites and partners. However, challenges remain in leveraging SEND datasets to their full capacity.

Sponsors may lack the infrastructure (eg resources, budget) to maintain a database and visualisation tools. Also, instituting that infrastructure often requires staff to change business processes to connect data science with operations, particularly when moving from fragmented, locally-owned data sources to centralised databases that offer richer data but more complexity. As a result, shared data that is available may be overanalysed or even misinterpreted.

In addition, there are key data omissions from the SEND standard that will limit its utility in terms of advanced analytics. For instance, there is no means to capture compound structure within the standard. In order to develop Quantitative Structure Activity Relationship (QSAR)-based models from the datasets, compound structure data must be linked with a SEND data warehouse.

Furthermore, there remain issues with the homogeneity by which key information is captured in the standard by different partners. There is arguably too great a flexibility in how the representation of study design, dose group identification and study elements are captured in the format. However, these issues could be overcome with a focused effort by stakeholders.

Precompetitive SEND use cases for safety/tox

Widespread use of the SEND standard will facilitate precompetitive collaborations by providing a standard format for sharing preclinical data and reducing the amount of harmonisation needed in order to create integrated datasets. The Innovative Medicines Initiative (IMI) project, Enhancing TRANslational SAFEty Assessment through Integrative Knowledge Management (eTRANSAFE)[5], aims to develop just such a dataset, utilising data from 12 pharmaceutical companies for translational safety analysis. In particular, it plans to address the following questions:

- What is the relevance of preclinical safety studies for clinical study safety assessment?
- Can we refine preclinical experiments to better predict clinical safety?
- Can we build predictive models of safety outcomes from the data?

By supplementing the SEND data with information on the structure and pharmacology of the drug, the dataset will support translational research analytics, such as determining the possible liabilities for new drug candidates or identifying mechanistic hypotheses for adverse effects.

As the amount of data in SEND format evolves, and by combining it with legacy data in the same format, it will be possible to identify the background incidence of histopathological findings along with the most common drug-related

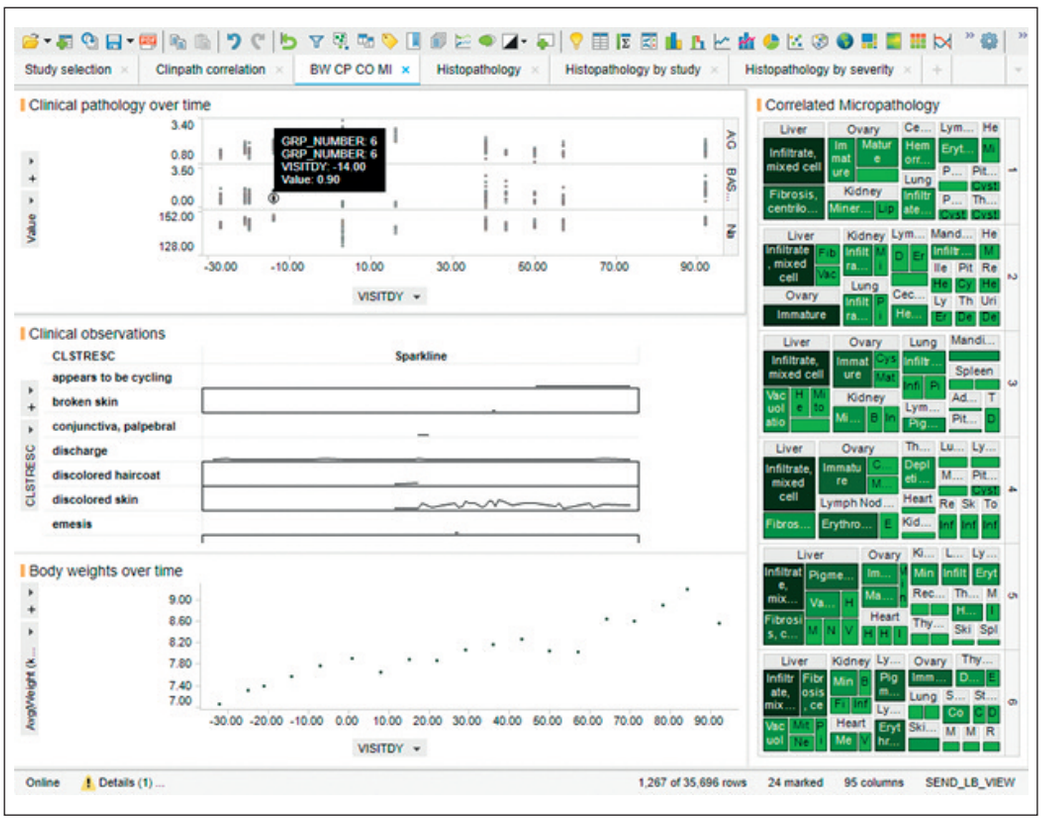


Figure 4
Supporting study monitoring through visualisation of clinical pathology, clinical observations, body weights and micropathology data

pathologies and determine reference ranges for clinical pathology. Drilling down further into this data can identify the impact of study duration, species specificity and gender differences on the findings observed.

Conclusion

Sponsors are now required to submit data from preclinical toxicology studies started after December 17, 2016 to the FDA in SEND format for Investigational New Drug (IND) applications and New Drug Applications (NDAs). The SEND format creates a structured terminology and consistency for regulatory submissions and allows for concise communication between the FDA, CROs and sponsors.

Moving beyond the objective of regulatory compliance, the SEND format standardisation represents a significant opportunity for drug developers to extract valuable scientific and operational insights from preclinical data. From enhanced study monitoring and visualisation of operational metrics to reducing drug failure rates and improving safety outcomes, the potential is significant at a time when risk and complexity of drug development continues to rise. Eventually these data sets can be leveraged to build predictive models that reduce the burden of *in vivo* testing and help identify/validate novel targets.

Ultimately, these efforts will complement systems biology approaches and serve to improve the drug development process, leading to better medicines being available for patients in a timelier fashion. DDW

Katherine Briggs is Research Leader at Lhasa Limited. Bob Friedman is Chief Technologist at Xybion Medical Systems. Kristen Ferrara is Director, Global R&D IT Programs, at Takeda Pharmaceuticals. Mark Pinches is Senior Principal Scientist at Lhasa Limited. Will Drewe is Principal Global Alliance Manager at Lhasa Limited. Cheryl Riel is Head, Nonclinical Writing and Document Management, at Alnylam Pharmaceuticals. Melody Thompson is Manager, Nonclinical Writing, at Alnylam Pharmaceuticals. Antoinette Hayes is Lead Associate Scientist, Toxicology Study Manager at Alnylam Pharmaceuticals. Jason Gratt is Principal Software Engineer at Takeda (formerly Millennium) Pharmaceuticals. Shylah Wyllie is Manager II, Drug Safety Research & Evaluation at Takeda Pharmaceuticals. MaryBeth Walsh is a consultant with the Pistoia Alliance.

References

1 CDISC-SEND. https://www.cdisc.org/standards/foundational/send.
2 Scope is INDs and NDAs Requirements are listed in the FDA data standards catalog, which can be found at: https://www.fda.gov/forindustry/datastandards/studydatastandards/default.htm.
3 Innovation in the pharmaceutical industry: New estimates of R&D costs. Journal of Health Economics 47 (2016) 20-33 https://dukespace.lib.duke.edu/dspace/bitstream/handle/10161/12742/DiMasi-Grabowski-Hansen-RnD-JHE-2016.pdf?sequence=1&isAllowed=y.
4 Figures 1-4 provided by Xybion Corporation.
5 The eTRANSAFE project (www.imi.europa.eu) has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 777365. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.