Robert King and George Weir
PPD Clinical Research Services, Thermo Fisher Scientific

Clinical Data Management:

Past, Present and Future

The ability to navigate the internet via a web search is a modern marvel, one that essentially enables the online lives to which we've grown accustomed. The beauty of it is you don't have to think about the number of databases involved or whatever different repositories are being crawled and catalogued on your behalf. You just enter your query and, more often than not, the search box returns an immediate, accurate response.

However, such ease of use is not yet the standard in clinical data management (CDM), where researchers and data managers must toil through coordinating databases and ensure all needed wseamless, Google-like search experience for clinical data review and cleaning is coming sooner than you may think—and getting closer every day.

The CDM field has seen many developments over the years, from data entry of 100% paper-based trials, to EDC (electronic data capture) becoming just another data source alongside eCOA (electronic clinical outcome assessment), central laboratory data, ECGs, etc., to the adoption of digital-enabled trials and the proliferation of data sources such as wearable devices, eSource, EHR (electronic health record) and biosensors. This gradual evolution has resulted in vestigial practices tied to older, paper-based models, such as "page" being the default term for collection modules in EDC; queries still being considered as a CDM responsibility versus being part of the overall data review/cleaning process; and study teams reviewing data listings that are commonly dumps of data to be manually and time-consumingly trawled through to

find anomalies and assess data quality and integrity.

Now, with technological advancements such as artificial intelligence (AI)/machine learning (ML) enabling us, CDM sees this very much as an exciting opportunity to drive ourselves into the digital age of real-time data collection and management. As the industry moves into this digital era that incorporates a much more patient-centric approach, the volume of data collected in clinical trials is going to increase exponentially, making it critical that CDM, as a function, finds ways to efficiently consolidate disparate data sources into one centralized place and ensure we can validate and review data according to existing demands. These demands include, but are not limited to:

Trial complexity: Today's clinical trial designs require real-time data modeling and simulation to support informative, data-driven decision-making and to reduce development time, costs, and late-stage research failures. For example, many clinical trials are now considered adaptive, meaning that they can change as the trial progresses and incoming data are used to dictate next steps. In such a scenario, if a patient is not reacting to a drug, the trial operators can decide to change that drug or its dosage. Immuno-oncology, multi-arm and collaboration trials also add new levels of complexity to clinical trials.

- **Master protocol designs:** Master protocols are designed with multiple substudies, which may have different objectives and involve coordinated efforts to evaluate one or more investiga-

tional drugs in one or more disease subtypes within the overall trial structure. These adaptive studies are generally classified as umbrella studies, basket studies and platform studies. Where one database build used to suffice, today's more complex protocols require flexible, dynamic database builds with multiple arms, cohorts and decision trees being incorporated to the EDC.

- Umbrella studies examine multiple drugs for one indication, with potentially different routes of administration that in turn increase complexities with data collection instrument design, investigational product (IP) distribution and safety reviews.

- Basket studies examine one drug for multiple indications, which requires increased domain expertise, multiple endpoints (usually one for each indication) and greater variation of the participant population (which also increases complexity for data reviews).

- Platform studies examine multiple drugs for multiple indications, thereby Inheriting the complexities of umbrella and basket designs, along with additional upfront planning to consider all data-related scenarios, the need for a greater number of interim analysis and the need for dedicated or specialized CDS teams to manage these multiple trials.

- **Digital enablement:** The primary challenge of digitally enabled EDC is to keep the data from simply becoming another siloed data source with limited access and usability. Such silos need to be replaced with a new "data-lake" model that houses all data from all sources.

- **Patient-centric focus:** Today's trial models are focused more on the needs of the patient rather than the clinical site. This includes a shift away from traditional EDC methods and a greater emphasis on direct capture of data from wearables and eCOA increases.

LESSONS FROM THE PAST
Gone are the days when standard line listings are waiting for you – if they have correctly run or if you've remembered to run them – at the bottom of a giant printer, with your username proudly showing in massive dot matrix characters at the top. Printed in a room that had to be separate from the main office due to the noise produced by the mammoth machine. Printed on paper with perforations at the bottom to make distribution to your colleagues easier.

An aspect of former times worth revisiting is having all patient-centric data in one centralized location. Accomplishing that in the past involved a raft of CDM data entry staff inputting data from paper CRFs, paper diaries, scales, lab reports, etc., into the database. But to be fair, it did result in having all patient data in one place to make its interrogation a little easier, albeit antiquated and very time-consuming compared to more recent years.

However, if we take ourselves back to those good old days, one could personally build one's own database, make one's own database amendments without reliance on database administra-

tors, do one's own programming and import data directly into the system. Coding of medical terminology was done yet again in the system. This one-person approach reduced the reliance on other functions to support the CDM effort.

In the past, clinical data managers/associates would be skilled in their reviews of the patient data by being able to perform reviews cross form, write SQL queries to pull outputs to check on validity and consistency of data points, review listings with a knowledge of the therapeutic area or indication to allow more structured queries to be raised back to the investigative site. This process, however, was heavily reliant on the skill and experience level of the person performing the reviews. Therefore, consistency of review and data validation was hard to achieve, explaining why inefficient and onerous QC steps had to be performed to give at least a sense of comfort of quality.

As nice as it may be to reminisce about those days in the late 1990s/early 2000s, let's bring ourselves back to today and the steps required for the industry to advance into the digital age in terms of data review and validation.

Today, although some of these same skills exist among clinical data managers/associates, these professionals can be compelled to look at data in silos. Thus, instead of looking and thinking about the eCRF being the story of the subject's navigation through a clinical trial, they think about it in terms of erroneous data points, don't look for trending, struggle with the concept of cross-form review and focus on edit checks to help them resolve a single point of entry without thinking much beyond that. What else could that impact? Does it affect previously locked data? Are there queries that exist elsewhere in the eCRF that pertain to the same issue? What about third-party vendor (TPV) data?

In our humble opinion, we have for some time been in the "dark ages" where not much has really changed in terms of reviews and how CDMs went about their business. Nevertheless, we shouldn't be too hard on ourselves, as we still managed to be very successful and deliver clean data to allow for decisions that brought medicines and treatments to much-needed areas.

Of course, some trailblazers dipped their toes into the water regarding visualizations and other dashboard tools helping to array data in a more visual way, but CDM never really got away from its edit check and listing processes of old to adopt these in any significant way. The lack of technological advance during these times kept CDM adhered to these standard processes, while we still maintained a very high level of data quality and integrity.

VISION OF THE FUTURE
With the introduction of novel technologies and new clinical procedures, we see more and more data being collected via digital means. The old ways of doing so (first paper and later EDC) are no longer practical options. With EDC becoming yet another data source, combined with the increased need to reduce database build times, data handling durations, increased productivity, all underpinned by a quality-by-design process to ensure data integrity, it's requiring clinical data management teams to:

- Embrace the opportunities that these technologies bring to automate data review, improve quality, and reduce manual burden and resource-intensive steps.

- Evolve from simply collecting, cleaning and providing data to internal/external customers to becoming data stewards and taking a leadership role within this rapidly changing digitally enabled environment.

- More than ever play a crucial role in delivering vastly more efficient processes for real-time data collection/processing, TPV integration and technology optimization, to ensure timely high-quality data delivery to internal/external customers.

- Further enhance the TPV selection and drive subsequent management processes to ensure key criteria are understood, agreed upon and adhered to.

- Ensure data alignment of TPV, on-time data delivery, seamless integration and reconciliation with other clinical data and improved delivery oversight.

The mechanisms and processes of capturing, reviewing and cleaning data need to be aligned. EDC platforms we currently have will not cope with such huge volumes of data, so repositories (such as a data lakes) are being introduced. Data managers and the clinical team will be unable to use the same data cleaning practices and instead are turning to AI/ML to scour the data initially, significantly improving the preliminary quality of the clinical data and highlighting potential errors for teams to interrogate further to determine if indeed the data is erroneous.

Through the implementation of advanced dynamic visualizations to extensively review the clinical data collected on a clinical trial, teams will identify trends and improve data quality and integrity much more readily than reviewing via "data dumps." Resulting in real-time review, reducing overall time to clean as outdated, time-consuming, retrospective data reviews will be replaced by a risk-based, prospective validation approach.

Additional enhancements to the data review process are to apply KRI (key risk indicator) thresholds that can focus the reviewer on areas of concern in the data or the flow/processing of the data. Study-specific KRI thresholds would be defined and introduced into the review process by programmatically adding them to the dashboards. Further moving the process of data cleaning away from a "review everything" mindset to a much more focused risk-based approach.

Continued enhancements and optimization of the "skills" of AI/ML, along with improvements in the use of KRI thresholds and technology enhancements, will further reduce the burden on teams even more. That too, will allow teams to increase their attention on critical thinking activities and making better informed decision faster, to deliver life-changing medicines.

To deliver on this exciting vision, CDM is beginning to embrace managing the end-to-end data life cycle through:

- Critical thinking of the requirements of the protocol design, key data, clinical endpoints, study setup and study conduct activities to assess risk to the trustworthiness of trial results.

- Advising on study design from a database design and data collection standpoint and providing guidance on the data collection strategies during study design.

- Manging both the internal and external interdependencies to effectively manage external vendor performance and data quality.

- Deploying AI/ML technologies (e.g., RPA) to automate repetitive and simple tasks, chatbots increasing study performance and data visibility, all improving productivity.

- Overseeing standards compliance across the end-to-end data life cycle.

CONCLUSION

Data collection, processing and reporting will continue to be more challenging in the future with the introduction of new disease areas, the growth of decentralized, patient-centric clinical trials; wearable usage becoming more common; the ever-increasing complexity of clinical protocols; and the exponential growth in the volume of data.

Clinical data management teams are central and critical to delivering future success. Data managers already are taking leadership roles in defining what the future will look like, from deploying and configuring systems, becoming data stewards, developing efficient processes and working with clinical development teams so they acquire new skills and apply a much more strategic approach to safeguarding data integrity. CP

ROBERT KING, Executive Director, Biometrics Functional Service Partnership Solutions, PPD Clinical Research Services, Thermo Fisher Scientific. Robert is an executive director with Thermo Fisher Scientific, where he is responsible for PPD Clinical Research Services' Biometrics Functional Service Partnership (FSP) Solutions, collaborating with clients to develop customized approaches to deliver resource continuity, increase productivity and drive value to bend the time and cost curve. Robert has over 30 years of experience in clinical development. During the last 15 years he has had global leadership roles in data management/biostatistics and clinical development support functions. In all these roles he has been a key driver in developing and implementing corporate strategic direction across clinical development and eClinical technologies, as well as gaining significant experience in the FSP marketplace.

GEORGE WEIR, Senior Director, Clinical Data Management, PPD Clinical Research Services, Thermo Fisher Scientific. George started his career with PPD Clinical Research Services in 1997 before leaving to take a position with pharma in 2000. George returned to PPD Clinical Research in 2005 where he has since held various management positions, most recently as senior director of clinical data management (CDM) with Thermo Fisher. With more than 24 years of CDM experience covering both cro and large pharma, George is home-based in Scotland where he currently provides oversight to all CDM FSP opportunities, ensuring adherence to timelines, budget, procedural documents