acdm

association for **clinical data management**

The implementation of the evolving Clinical Data Science role, a cross collaborative initiative

Authors: Tanya du Plessis, Hari Priya, Rashida Rampurawala, Rich Davies, Vishal Kapoor, Eva Gjerlevsen Harreskov, Sivakumar Pazhamalai, Andrew Owain Green, Santosh Karthikeyan Viswanathan

Contributors: Akhila Velicheti, Magdalena Wozniak, Linda Shostak, Peter Sec, Alberto Clemente, Lauren Gray

The implementation of the evolving Clinical Data Science role, a cross collaborative initiative

One of the most recent topics in the industry is the introduction of data science to clinical trial strategy, execution and evaluation. Although adoption approaches differ based on individual companies' data strategy, the reasons for the drive behind the change is clear, and these include:

- Data sources which have become more and more diverse. In the past clinical trials were very linear in data collection and mainly centralized around the EDC (Electronic Data collection) platform. Today there are multiple different data sources (most of these hosting eSource data) holding the same if not more data than the EDC platform
- The way data is now being collected through wearables and other eCOA driven methods, as well as huge laboratory datasets such as genomic datasets, produced more data points than we have ever seen before
- The ability to collate all data sources in order to make insightful and timely decisions is not only a necessity to ensuring patient safety but also a cost saver in ensuring decisions are being made at the right time. This drives the need to have data flowing into a collated space and much higher frequencies than we have managed in the past
- Recent regulation updates (specifically ICH E8 and ICH E6) which call for data surveillance and a risk strategy inclusive of data orientated risks
- The release of ML and AI tools to not just expedite processes but also allow for much higher volumes of data processing. This is already common practice in many other industries and there is an external pressure to ensure we are using this technology to ensure the progression of the pharmaceutical industry too
- Then lastly, there has been a realization in the industry that the data collected from the clinical trials should be seen as an asset to be mined and evaluated for insights to support future trials. This not only mitigates risks for future studies but also can help in the strategic planning of those trials.

From the above requirements, tools and/or technologies are required to facilitate these requirements. There are of course many technologies available with a range of abilities, however at minimum a company's' strategy should include a data collation environment (eg: A CDR/MDR or enterprise data lake or something similar) and an analytics (preferably including visualizations) platform connected to the collated and ingested data.

The specific focus area for a company is dependent on several variables such as Therapeutic area (and/or focus indication), general trial designs, company size, historical areas of focus, etc… and this additionally seems to be driving in which group (or groups) within a company that is driving the change.

Although this requires a holistic study team approach, the need for additional insights into the data is a major driver.

The implementation of the evolving Clinical Data Science role, a cross collaborative initiative

1. What is the difference between "clinical data science" and "data science" ?

A diversified approach is evident where there seem to be a few different approaches to address the need to explore data science with organizations today. For some this is being addressed through the creation of new and different roles, or alternatively by adding on to the responsibilities of current roles. Although the exact roles and the departments in which these roles function have yet to be fully established with in the industry, there does seem to be two main thought lines and a high-level split between companies who are defining approaches between "clinical data science" and "data science".

Within these definitions the major differentiator between "clinical data science" and "data science" seems to be the roles of the team members within these teams, as well as which department the team is housed in. Clinical data science is generally housed within the data management operational team whereas the Data science team within a departmental team (such as statistics or DM), or part of the IT team; or even within an individual team. Data science may have broader responsibilities in supporting the entire business.

The general differences between these roles are defined below on teams working with the clinical trial data, however it should not be overlooked that there are hybrid versions of these being applied within some organizations.

Data Scientist	Clinical Data Scientist
Collecting & organising large datasets	Managing the collection of datasets
Not trial specific data	Clinical trial data
Cleaning & pre-processing data	Developing and implementing data cleaning & pre-processing strategy
Statistical and machine learning techniques	Interpretation of the outputs from statistical & machine learning
Building predictive models for decisions & solutions	Translation into actions & strategies
Stakeholders across the organization, including executives, product managers, & engineers	Variety of stakeholders, including study team
Within a departmental team (such as statistics or DM), or part of the IT team	Data management operational team
← Diverse datasets from various sources →	
← Data quality & integrity →	
← Identify patterns and relationships within the data →	
← Data insights →	
← Communicating complex data analyses and findings →	

1.1 Current definition and Job description of a Data scientist:

Data scientists typically use statistical and machine learning methods to build predictive models, and seem to be generally trial agnostic in their approaches. Their main purpose being to create models that can analyze large datasets to identify trends, patterns, and insights. They are, however, not often part of the team that is actually taking action on risks, trends or results found, this is typically handed over to the study team and business teams.

The job description can vary depending on the organization however some of the common responsibilities may include:

- **Collecting and organizing large datasets from various sources, including databases and APIs**
- **Cleaning and preprocessing the data to ensure accuracy, completeness, and consistency**
- **Applying statistical and machine learning techniques to identify patterns and relationships within the data**
- **Building predictive models to make data-driven decisions and develop innovative solutions to business problems**
- **Communicating complex data analyses and findings to stakeholders across the organization, including executives, product managers, and engineers.**

The qualifications of this team are typically a graduate degree (such as a Bachelor's, Master's or PhD) in a data science, statistics, or biostatistics. They need to have some hands-on experience in statistical software such as R, SAS, or SPSS, as well as programming languages such as Python or SQL.

Overall, data scientists play a critical role in using data to drive business decisions and improve organizational performance. They leverage their skills and expertise in statistics, programming, and machine learning to extract insights from data and transform it into actionable knowledge.

1.2 Current definition and Job description of a Clinical data scientist:

Clinical data scientists, on the other hand, work with data directly from the clinical trial they are specifically working on. Responsibilities can also overlap with some companies' centralized monitoring team they are often primarily responsible for management of data related risks for a clinical trial. This may include the identification, mitigation strategies, monitoring and continual assessment of risks to the study data which may impact patient safety or study end points.

While both data scientists and clinical data scientists have a strong foundation in understanding the regulatory requirements that govern the collection, storage, and analysis of healthcare data; a clinical data scientists' focus will typically be directed to a specific trial and the associated risks.

They are highly skilled at interpreting data generated from clinical trials, electronic health records, and other sources of medical data. They may work in collaboration with a data scientist to use advanced statistical and computational methods to extract insights and trends from these datasets. Or, in some companies perform this task independently though the use of technologies and tools such as data driven risk evaluation platforms, or analytics and/or visualization platforms.

The job description can vary depending on the organization however some of the common responsibilities may include:

- Managing the collection of datasets from various sources for the clinical trial
- Developing and implementing data cleaning and pre-processing strategy to ensure data quality and integrity
- Interpretation of the outputs from statistical and machine learning methods to identify patterns and relationships within the data
- Collaborating with the study team to translate data insights into actions and strategies
- Communicating data analyses and findings to a variety of stakeholders, including study team.

The qualifications of this team are typically a graduate degree in a related field such as medical science, natural science, data science or statistics; and many companies accept a Bachelor's degree with relevant experience. It is however critical that this team has extensive knowledge in the execution of clinical trials from a data perspective.

Overall, a clinical data scientist plays a crucial role in management of data strategies, risk identification and management, as well as driving the actions to be taken from the insights obtained from the clinical trial data. They leverage their skills and expertise in the understanding of clinical trial execution, clinical trial data, and therapeutic areas to extract insights from data and drive data integrity.

2. Team structures and growth

When thinking about the management of clinical trial data it's critically important for a company to have a clear vision for their biometrics department as to why they are making a change; and what the organization wants to (or needs to achieve) to achieve this; in order to set achievable goals. It must be clearly communicated throughout the entire organization to ensure there is understanding and support for the change.

Above we discuss the differences between Data science and Clinical Data science, however it is important to remember that this approach is heavily intertwined with clinical operational activities as well. Therefore, there are currently numerous organizational structural approaches to housing these new roles/activities...

1) Creating a new team

- Completely new team/department within an existing operational team. This new team consists of team members which come from existing departments with the new responsibility of driving the risk management process and analytics. They supplement the current operational team structure/processes.

2) Adding new roles to departments.

- The addition of new roles to supplement the current team experience
  - For example: Data Management adding a Clinical Data Scientist role, and/or IT adding a Data Scientist role
  - For example: Centralize monitoring changing the full scope and adding a Clinical Data Scientist role, and/or a Data Scientist role.

At this point it important to note that for smaller organizations such changes are just simply not possible due to constrained head count. This may mean that the role/s and responsibilities may need to be added to the job description of an existing role.

No matter the size of the organization nor the structural approach, it is critically important that the entire study team are aware of the expectations and process changes. It must be well explained how these new roles/responsibilities/processes will be embedded within current processes to ensure that the team understand how to work together and when responsibility overlaps.

It can also not go unsaid that technology and tools have a direct impact on the processes planned or being followed. If an organization has a specific technology or tool in mind as part of their technology roadmap, then it is important to understand how this could impact the planned operational changes before finalizing the plan.

3. How do we grow/develop our teams today:

Regardless of whether a company is large enough to make structural changes, or if they will simply be adding responsibilities to existing roles, every company will need to perform a gap analysis to determine the exact gap of skills and expertise for their needs within their organization specifically. The success of this approach, however, is determined by how clearly the expectations are set around the new roles, or skills sets, required. In order to get this right from the beginning the organization must have a well-defined vision of what they want to (or needs to achieve) in order to set an achievable goal. The result of this would typically be to upskill/cross skill current employees and/or appoint new skills for larger gaps.

Upskilling can be achieved through various means, including but not limited to....

- Training (Example: LinkedIn learning, Udemy, etc...)
- Industry education/awareness (Example: organizations such as ACDM)

- Reverse mentoring (Bottom-up training: Employees with less years industry experience but more years in data science/analytics to train upwards in the organization)
- Job rotation (Employees are encouraged to apply for different roles for a "Limited time" to gain better exposure and understanding of other roles)
- Train the trainer sessions (Community of Practice sessions on specific topics like knowledge graph, generative AI etc)
- Tech vendor support (Technology vendors will help guide through the adoption process and provide insights on what to expect).

The question, however, often comes up as to when the right time is to rather look outwards and recruit the expertise needed. Examples of when this should be considered are…

- Gap analysis shows specific gaps within organization that cannot simply be covered by upskilling existing staff, then recruit to fill gaps (An example will be if an organization would like to take a data science approach, then additional programming and technology skills are required through a qualification)
- Introduction of new tools and technologies may require different skill sets.

It is also common practice currently to approach this transformation in stages rather than trying to address all gaps at once. Success around this approach includes limiting upfront expenses (staggering these across a few month/years) and allowing all study team members enough time to prepare and understand the changes in phases rather than overwhelming teams with a lot of new information.

4. Metrics to measure success of implementation (quality)

Finally, it is very important to ensure that the success of the implementation of the changes to meet the industry demands are being measured. Since the specifics of these are defined by an organization planned goals, it is difficult to provide exact measurements or Key performance indicators (KPI's) to be used. However general KPIs/goals could include…

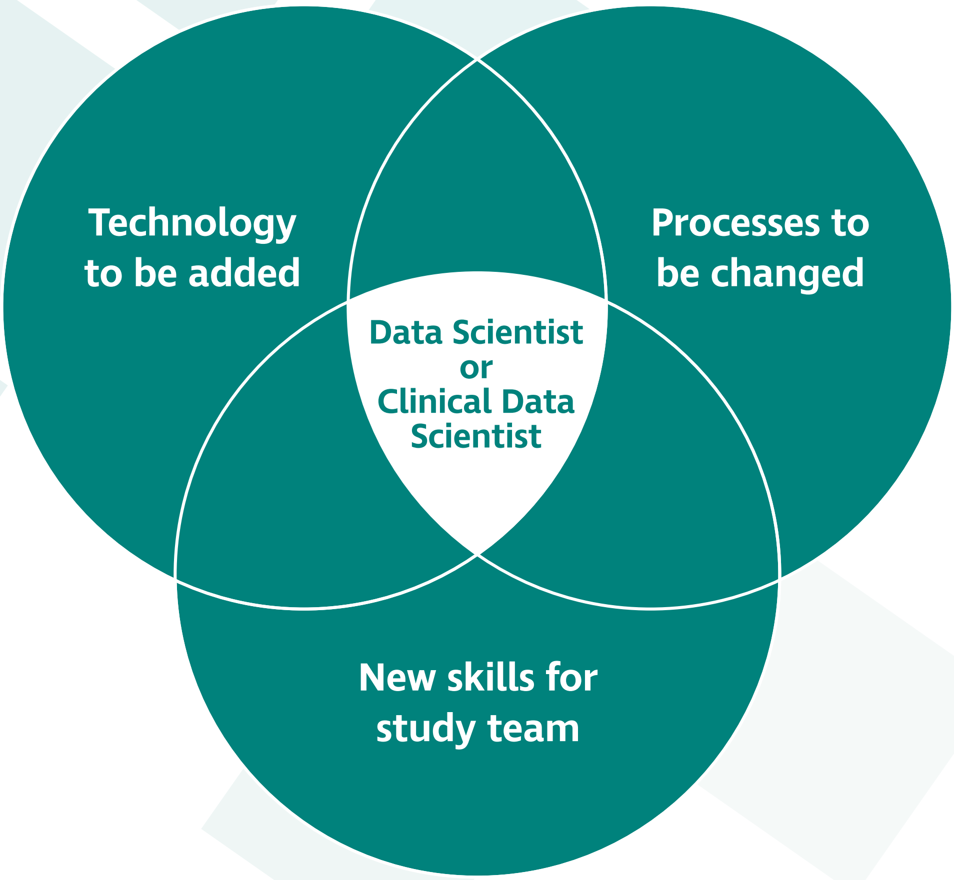
- Risk evaluation to include "X" number of data focused risks at a minimum
- The ability to identify data issues through unsupervised data surveillance techniques, that would have been missed or found too far down the chain of events to be addressed in a timely manner. Track, how, when and where it was found
- Time from data issue entry/event occurrence to identification
- Time from issue identification to decision made
- Accelerated Clinical trials (Getting data in and evaluated quicker)
- Quality metrics in relation to study data risks.

5. Conclusion and summary

In conclusion, although many organizations have already taken the leap and implemented process, operational and structural changes to adapt to the industry requirements, many companies are still in the process of adopting or planning out their approach to clinical data science and/or data science.

The structural changes seem to be the most difficult to navigate at this stage because the change impacts the entire study team, and there are still many questions unanswered in regard to where new role/s fit best. There is currently a very distinct difference in how the role of Clinical Data scientist and Data Scientist are defined within companies, in some companies both apply. However, since many companies are still in the process of defining this, when working with someone with this title it is recommended to confirm what their responsibilities are and not to assume.

Importance of managing change



In order to address the changes needed there is a balance to be found between technology to be added (with many well experienced tools/technologies already available); processes to be changed (adapting to the technology and practical challenges of data changes) and finally the study team to be complimented with new skills (either through addition of new roles added, or upskilling of current team). Throughout this process it is very important to manage these changes within an organization through clear communication around the vision to be achieved and the goals being set.

acdm

association for **clinical data management**

www.acdmglobal.org