Paper PD01

Statistical Programming 101 for the Fourth Industrial Revolution

Ritika Aggarwal, Novartis Healthcare Pvt. Ltd., Hyderabad, India

ABSTRACT

The Fourth Industrial Revolution is the advent of "cyber-physical systems" involving new capabilities for people and machines. It builds on the fusion of technologies from physical, digital, and biological spheres. According to a survey conducted by the Economic Intelligence Unit, the healthcare sector would gain the most from the merging of these systems. Consequently, the method for collecting comprehensive health and fitness data is gradually making headway. For example, Novartis developed a digital inhaler that could enable COPD patients to observe data on their own inhalation use in real time. Inevitably, the constantly shifting complexity and diversity of healthcare data will cause major disruptions and surprises for the statistical programming function of the pharmaceutical "industry 4.0". So, how can the statistical programmers prepare for the imminent challenges and identify sustainable solutions?

For a forearmed future of statistical programming, this paper offers a few potential solutions –

- Understanding end-to-end data lineage
- Being fully cognizant of stakeholder's expectations and rationale.
- Getting ready for the future of statistics and data science.
- Leading an agile team of programmers

INTRODUCTION

The fourth industrial revolution (4IR) is characterized by the fusion of digital technologies, artificial intelligence (AI), the Internet of Things (IoT), and other advanced technologies, which are transforming various industries. In the pharmaceutical industry, the 4IR is bringing about significant changes and innovations. It is reshaping the industry by leveraging these technologies to accelerate drug discovery, improve patient care, and enhance the efficiency of manufacturing and supply chain processes. Data lies at the heart of transformation in Pharma 4.0 (introduced in 2017 by ISPE). Data is integral to the entire lifecycle of pharmaceutical products, from drug discovery and development to manufacturing, regulatory compliance, and post-market surveillance. The industry's ability to harness and analyze diverse datasets is key to advancing scientific knowledge and driving innovation.

The Pharma 4.0's statistical programming function is essential in analyzing and interpreting data derived from many sources, such as clinical trials, observational research, and real-world data. It is necessary to transform raw data into meaningful insights that are important for making decisions, meeting regulatory requirements, and advancing drug development and safety. The 4IR has had a profound impact on data collection, leading to significant changes in how data is gathered, processed and utilized. There has been a surge in IoT and sensor technologies, big data analytics, machine learning and AI, edge computing, augmented and virtual reality etc. The collection and storage of clinical trial data are shifting from paper-based to electronic-based methods. Moreover, various clinical trial management systems (CTMS) and clinical data management systems (CDMS) are already in use for capturing, storing, managing, and reporting clinical trial data in real time that can be submitted to the health authorities. Inevitably, the shifting complexity and diversity of clinical data will significantly impact one of the major consumers of data and metadata in Pharma 4.0 – the Statistical Programmers. As the pharmaceutical industry embraces the 4IR, statistical programmers play a crucial role in ensuring that data analysis meets regulatory standards and supports drug development based on new technology. Therefore, it becomes essential to examine our digital maturity, current technological demands, and a transformation roadmap to stay up to date with faster-than-ever speed of change toward a hopeful future of the industry and individual development. This paper focuses on a few potential challenges for the statistical programming function and suggests some individual and group-level strategies to address them.

Byte-sized Battles: Statistical Programmers Tackling Challenges in Pharma 4.0

The integration of advanced technologies and the evolving nature of data in the industry bring about unique hurdles. Here are some challenges faced by statistical programmers in Pharma 4.0:

Complexity of Clinical Trial Data

The fourth industrial revolution has brought about an explosion of data, including electronic health records, genomics data, wearable device data, and more. Clinical trials now often deal with large datasets, requiring advanced analytics and complex statistical models to derive meaningful insights. Traditional clinical trial endpoints are being supplemented or replaced by digital endpoints, which involve the use of data from wearable devices, mobile apps, and other digital sources. For example, Stride velocity 95th centile (SV95C) is the first wearable acquired digital endpoint to receive a qualification from the European Medicines Agency (EMA) to quantify the ambulation ability of ambulant DMD patients aged ≥5 years in drug therapeutic studies (Servais L et al., 2021). Remote monitoring allows for real-time data collection, enabling a more dynamic and comprehensive understanding of a participant's health. However, integrating RWE into clinical trials adds complexity in terms of data variability and quality. Moreover, with the increased volume and diversity of data, it becomes important to evaluate interoperability and data standardization while maintaining data security (Popov VV et al., 2022). Ensuring that different systems and devices can communicate seamlessly, and that data adhere to common standards is essential for efficient and meaningful analysis. To adapt to the changing complexity of the data and technology, a statistical programmer of Pharma 4.0 must know how to:

- Navigate through heterogeneous datasets, varying data formats and standards. Account for data variability and integrate data accordingly.
- Understand the science behind the product/molecule to understand data better.
- Stay updated on modern technology and processes to apply them appropriately to clinical data and derive meaningful insights from it efficiently.

Regulatory Compliance and Stakeholder Management

Regulatory bodies are adapting to the changing landscape of clinical trial data. There is an ongoing effort to establish guidelines and standards that accommodate the use of advanced technologies while ensuring data integrity, patient safety, and ethical conduct of trials. A significant step in this arena was the U.S. Food and Drug Administration (FDA) 2021 draft guidance on “Digital Health Technologies for Remote Data Acquisition in Clinical Investigations”. Statistical programmers in the fourth industrial revolution encounter specific challenges related to regulatory compliance and meeting the expectations of various stakeholders. The role of statistical programmers becomes increasingly crucial in ensuring that analyses adhere to regulatory standards and align with the expectations of diverse stakeholders. Here are some things that are needed to support regulatory compliance and improve stakeholder management:

- Regulatory agencies often emphasize the importance of interoperability and data standardization to ensure data quality and consistency. Statistical programmers may face challenges in integrating data from diverse sources while adhering to standardized formats and maintaining data integrity.
- There is an increasing expectation for transparency in clinical trial analyses. Regulatory agencies and other stakeholders often demand that statistical programmers provide clear documentation, code transparency, and methodologies that support the reproducibility of results. Achieving transparency is vital for building trust in the research findings.
- Effectively communicating with regulatory authorities is a skill that statistical programmers need to develop. Clear and concise documentation of statistical methods, data handling, and results interpretation is essential for regulatory submissions. The ability to respond to regulatory queries and provide justifications for analytical choices is crucial.
- As clinical trials become more global, statistical programmers face challenges related to the harmonization of international standards. Different regulatory authorities may have varying expectations, and programmers need to navigate these differences to ensure global acceptance of trial results.
- Statistical programmers need to align their analyses with the expectations of diverse stakeholders, including regulatory authorities, sponsors or vendors, clinicians, and patient advocacy groups. Balancing these expectations while maintaining scientific rigor is a challenging aspect of their role.

Skilled workforce

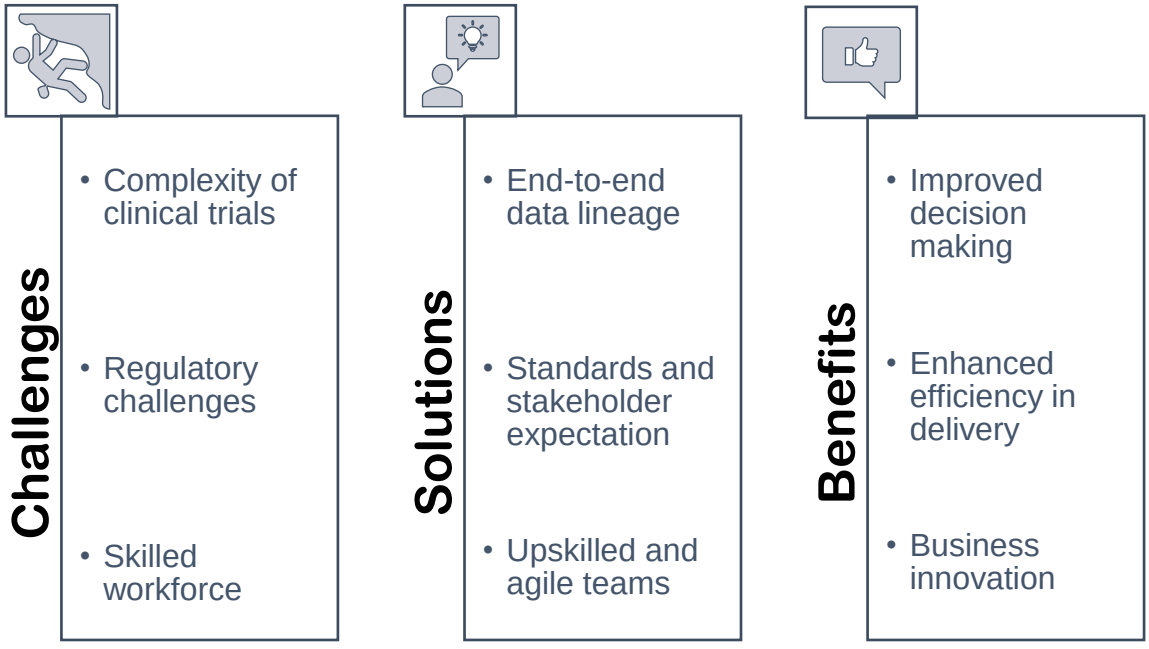
A skilled statistical programming workforce is critical for navigating the complexities of the fourth industrial revolution in the pharmaceutical industry, as discussed so far in this paper. Continuous learning and upskilling are crucial to stay

relevant in the rapidly evolving landscape of the fourth industrial revolution. However, there are several challenges that statistical programmers may face in their pursuit of ongoing education and skill development:

- **Rapid Technological Advancements and evolving regulatory guidelines:** In the fourth industrial revolution, new technologies and tools emerge at a rapid pace. Keeping up with the latest advancements in statistical methodologies, programming languages, data science tools, and regulatory standards can be challenging, requiring consistent effort to stay informed.
- **Diverse Skill Set Requirements:** Statistical programmers need a diverse skill set that includes proficiency in statistical methods, programming languages, data visualization, domain-specific knowledge, and people management. Balancing and continually enhancing this diverse skill set can be challenging, as it requires time and effort across multiple areas.
- **Time Constraints:** Statistical programmers often work under tight deadlines in the context of clinical trials and research projects. Finding time for continuous learning amid project demands can be challenging. Striking a balance between project responsibilities and educational pursuits is a common challenge.

Redefining Possibilities: Solutions and Strategic Next Moves

Addressing these challenges requires a proactive approach to professional development. It involves setting clear learning goals, prioritizing areas of improvement, seeking mentorship, participating in communities of practice, and leveraging a combination of formal and informal learning opportunities. A proactive approach includes strategic and well-rounded effort by both individual statistical programmers and their employers. It is a dynamic process that adapts to industry changes, technological advancements, and the evolving needs of pharmaceutical research and development.



End-to-end Data Lineage

End-to-end data lineage is a critical concept in the context of statistical programming in clinical trials. It refers to the ability to track and understand the flow of data from its origin (data source) to its final destination (analysis results). It helps in providing a comprehensive view of how data is transformed and processed throughout its lifecycle. This is particularly important in clinical trials where data accuracy, integrity, and traceability are paramount. Here is a breakdown of the key components and considerations for end-to-end data lineage in clinical trials:

1. Data Source Identification and Integration:
    - Identify and understand all data sources involved in the clinical trial, such as electronic health records (EHRs), case report forms (CRFs), laboratory data, and any other relevant sources.
    - Identify the data elements and variables within each source with the help of CDISC standards or data transfer specifications in case of third-party data.
    - Track data collection and integration from various sources into a centralized data repository.
    - Clearly understand data transformation or reconciliation performed by the cross-functional teams.

2. Data Transformation into analysis-ready datasets:
    - Learn to align data with study objectives/endpoints following defined standards.
    - Properly document the programming logics, or tools used for data manipulation and transformation. Streamline the process by following standard procedure to avoid any discrepancies.
    - Implement a robust metadata management system to streamline data transformation, bring uniformity in analysis and avoid potential bias.

3. Analysis and documentation
    - Ensure transparency in the handling of missing data, outliers, and any data imputation techniques by documenting the concepts as per the study requirements.
    - Version controlled programs, data and deliverables to facilitate reproducibility and auditing.
    - Annotation of the study protocols, statistical analysis plans (SAPs), and other relevant documents to map data and deliverables back to the study requirements. Similarly, establish mechanisms for traceability to link analysis results back to their source data.
    - Implement validation checks and quality assurance procedures at each stage of the data processing and analysis pipeline.

Regulatory Standards and Stakeholder Expectations

Adhering to evolving regulatory standards is a perpetual challenge. To make sure that the statistical analyses follow regulatory standards, statistical programmers must stay updated with changing regulations and guidelines. Along with a proper understanding of data, it is also important to be able to communicate the specifications of the work from a programming standpoint. The statistical programming function of the pharmaceutical industry is the critical cog in the clinical trial submission process. In this process, programmers actively communicate and collaborate with statisticians, data managers, and clinicians. For the submission process to go smoothly, the programmers need to clearly understand the upstream and downstream expectations. The following is what statistical programmers must do to ensure their work meets the regulatory requirements and stakeholder expectations, which will result in shorter turnaround times:

1. A thorough understanding of the statistical analysis plan and rationale behind the study objectives and the related deliverables will help the programmers contribute effectively to identifying relevant deliverables and setting up relevant analysis data.
2. Learning how statisticians and clinicians perceive data in alignment with the study objectives will help the programmers review their analysis effectively. This will reduce delivery time and increase the quality of the deliverables.
3. Knowledge of stakeholder expectations and regulatory guidelines will help in keeping the analysis (codes and scripts) and processes transparent, audit-ready, and reliable from the statistical programming function.

By incorporating these principles and practices into the statistical programming workflow, clinical trial teams can enhance the transparency, reproducibility, and reliability of their analyses, ultimately contributing to the integrity of clinical trial data and the validity of study findings.

Upskilled and agile teams

As discussed in the paper so far, statistical programmers play a key role in optimizing trial designs, implementing efficient statistical methodologies, and streamlining data analysis processes, contributing to faster and more cost-effective drug development. With the advent of the fourth industrial revolution, there has been a massive and continuous change in the skill and knowledge landscape. It is imperative that statistical programmers have a holistic understanding

of their roles in the industry and move on from the traditional ways of working. To thrive in Pharma 4.0, statistical programmers must opt for a comprehensive approach that addresses technical skills, methodologies, and stakeholder engagement. Addressing the challenges faced by the programmers in the context of Pharma 4.0 necessitates a proactive and strategic approach to professional development. Here's a more detailed exploration of the key components:

1. **Setting Clear Learning Goals:** Statistical programmers should identify and articulate their specific learning objectives. These goals may include acquiring proficiency in new programming languages, mastering advanced statistical techniques, or gaining expertise in the use of specific tools and technologies relevant to Pharma 4.0.

2. **Prioritizing Areas of Improvement:** Understanding one's strengths and weaknesses is crucial. Statistical programmers should prioritize areas where improvement is needed the most. This could involve honing skills related to data management, data visualization, machine learning, or specific regulatory requirements.

3. **Seeking Mentorship:** Mentorship is a valuable resource for professional development. Experienced mentors can provide guidance, share insights, and offer practical advice based on their own experiences. Mentorship relationships can help statistical programmers navigate challenges, enhance their skills, and gain a broader perspective on their roles.

4. **Participating in Communities of Practice:** Joining communities of practice within the pharmaceutical and statistical programming domains provides a platform for networking and knowledge exchange. These communities may include online forums, industry conferences, and local meetups. Engaging in discussions with peers facilitates the sharing of best practices, emerging trends, and solutions to common challenges.

5. **Leveraging Formal and Informal Learning Opportunities:** Professional development should encompass a mix of formal and informal learning. Formal opportunities may include workshops, courses, certifications, and degree programs. Informal learning can involve self-directed study, webinars, podcasts, and hands-on projects. This combination ensures a well-rounded skill set and keeps professionals adaptable in a rapidly changing industry.

6. **Employer Support and Culture:** Employers play a crucial role in fostering a culture of continuous learning. Organizations should encourage employees to dedicate time to professional development, provide access to relevant resources, and support participation in training programs and conferences. Creating an environment where learning is valued contributes to the overall growth and effectiveness of the statistical programming team.

7. **Encouraging a Growth Mindset:** A growth mindset is essential for navigating challenges and embracing learning opportunities. Both individuals and organizations should recognize that skills can be developed over time through effort and dedication. Encouraging a growth mindset fosters resilience, innovation, and a commitment to continuous improvement.

8. **Monitoring and Evaluating Progress:** Regularly assessing progress toward learning goals is essential. Statistical programmers should engage in self-assessment, seeking feedback from mentors and peers. Employers can facilitate this process through performance evaluations, recognizing achievements, and providing constructive feedback to guide further development.

The Evolution Blueprint: Crafting Solutions, Paving the Next Steps

The fourth industrial revolution has introduced both opportunities and challenges in the realm of clinical trial data. Embracing technological advancements while addressing issues related to data complexity, interoperability, privacy, and security is crucial for advancing the efficiency and effectiveness of clinical trials in this era. Even though these rapid advances are abstract, the good news is that the evolution of the fourth industrial revolution is within our power. Everyone should have a say in how technology affects them. Moreover, all industrial revolutions in the history of

humankind have fueled massive growth in how we perceive our work and society. Therefore, the key is to shape the fourth industrial revolution in a way that helps build a sustainable career that aligns with the future of work.

Continuous learning and adaptation are key elements to ensure ongoing success in the rapidly evolving landscape of the fourth industrial revolution. This paper suggests two models for statistical programmers to thrive in the same:

A learning model *that aims to equip statistical programmers with a holistic skill set (on an individual level), encompassing technical proficiency, agile methodologies, understanding of data lineage, and effective stakeholder engagement.*

Learning Model for Statistical Programmers in the Fourth Industrial Revolution

1. **Foundational Statistical and Programming Skills**:

Objective: Develop a solid foundation in statistical methodologies and programming languages commonly used in the pharmaceutical industry (e.g., SAS, R, Python).
Learning Activities:
- Gain proficiency in advanced statistical methods.
- Hands-on coding exercises and projects to apply statistical concepts.
- Participation in coding communities and forums.

2. **Understanding End-to-End Data Lineage:**

Objective: Develop the ability to understand, document, and manage end-to-end data lineage in the context of clinical trials.
Learning Activities:
- Map data flows and document data lineage.
- Thorough understanding of data management and data governance.
- Exposure to tools for metadata management and data lineage tracking.
- Case studies and projects involving the integration of diverse data sources.

3. **Agile Methodology and Project Management:**

Objective: Acquire skills in agile methodologies to enhance project management efficiency.
Learning Activities:
- Learn about Agile and Scrum methodology.
- Participation in agile-driven projects or simulations.
- Collaboration with cross-functional teams to implement agile practices.

4. **Regulatory Compliance:**

Objective: Stay abreast of evolving regulatory requirements and compliance standards.
Learning Activities:
- Stay up to date with regulatory and data standards like CDISC standards.
- Aim to learn about the submission process and requirements for various health authorities.
- Follow Good Clinical and Programming Practices.

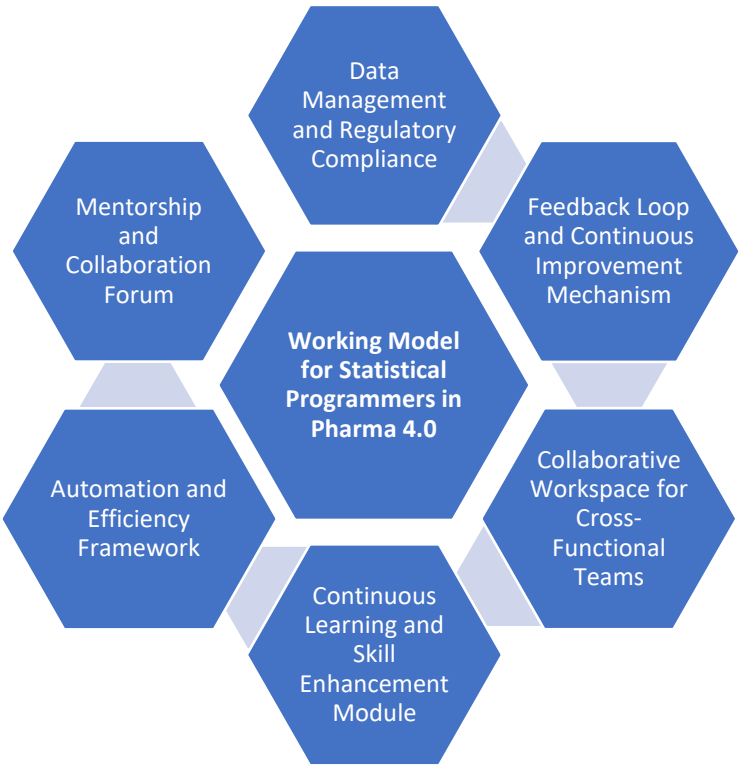
5. **Effective Communication and Stakeholder Engagement:**

Objective: Develop effective communication and collaboration skills to meet stakeholder expectations.
Learning Activities:
- Communication and presentation skills workshops – case studies based on project work.
- Collaboration in cross-functional teams on projects.
- Simulated scenarios for responding to stakeholder queries or building the ability to understand the rationale behind their requirements.

6. **Continuous Learning and Adaptation:**

Objective: Cultivate a mindset of continuous learning and adaptability.
Learning Activities:
- Regular participation in conferences, webinars, and industry events.
- Subscription to relevant journals and publications.
- Mentorship programs and networking with industry professionals.
- Understand and adhere to ethical considerations in statistical programming.
- Continuous feedback loops for improvement.
- Promote a culture of knowledge sharing and collaborative learning.

A working model *for statistical programming function in Pharma 4.0 (on a team level) that involves incorporating key elements that address the challenges and opportunities of this technological era. This model aims at redesigning or refining working practices for the statistical programming (and related) function which will require involvement of the leadership teams and collaboration with the appropriate stakeholders.*



Working Model for Statistical Programming function in the Fourth Industrial Revolution

1. **Data Management and Regulatory Compliance:**

Objective: Establish a centralized platform for managing and analyzing diverse datasets in adherence to Pharma 4.0 standards.
- Integration with advanced statistical software/programming languages.
- Automated data cleaning and preprocessing pipelines.
- Real-time collaboration features for cross-functional teams.
- Ensure adherence to evolving regulatory requirements and maintain data integrity

2. **Feedback Loop and Continuous Improvement Mechanism:**

Objective: Establish a feedback loop for ongoing improvement.
- Regular surveys for participant feedback.
- Iterative improvement meetings.
- Agile development approach for model enhancements.

3. **Collaborative Workspace for Cross-Functional Teams:**

Objective: Facilitate seamless collaboration among statistical programmers, data scientists, statisticians, clinicians, and regulatory professionals.
- Shared project dashboards.
- Integration with project management tools.

4. **Continuous Learning and Skill Enhancement Module:**

Objective: Foster a culture of continuous learning to keep statistical programmers abreast of industry developments.
- Access to online courses and workshops.
- Regular knowledge-sharing sessions.
- Integration with a learning management system (LMS).

5. **Automation and Efficiency Framework:**

Objective: Implement automation to streamline statistical programming workflows and improve efficiency.
- Automated code generation.
- Workflow optimization tools.
- Integration with version control systems.
- Interactive dashboards with key performance indicators.
- Monitoring of project milestones.
- Alerts for potential issues.

6. **Mentorship and Collaboration Forum:**

Objective: Facilitate mentorship and collaboration among statistical programmers.
- Mentorship matching platform.
- Discussion forums for knowledge exchange.
- Periodic networking events.

These models aim to provide a plan of action to the professionals on both individual and team levels. The best learning strategy is to shape the process in a way that helps individual professional development and nourishes effective team collaboration.

CONCLUSION

This paper discusses the impact of the fourth industrial revolution (4IR) on the pharmaceutical industry, specifically focusing on the role of statistical programmers in Pharma 4.0. The 4IR involves the integration of digital technologies, AI, IoT, and other advanced technologies, bringing significant changes to drug discovery, patient care, and manufacturing processes. Data plays a central role in Pharma 4.0, and statistical programmers are essential for analyzing and interpreting diverse datasets from clinical trials and real-world sources.

The challenges faced by statistical programmers in Pharma 4.0 include the complexity of clinical trial data, regulatory compliance, and the need for a skilled workforce. The evolving nature of data, advancements in technology, and globalized clinical trials add to the complexity. The text emphasizes the importance of staying updated on regulatory requirements, ensuring data standardization, and effective communication with stakeholders.

To address these challenges, the paper proposes a learning model for individual statistical programmers, emphasizing foundational statistical and programming skills, understanding end-to-end data lineage, adopting agile methodologies, staying compliant with regulations, and developing effective communication skills. Additionally, a working model for the statistical programming function is suggested, focusing on data management, continuous improvement, collaboration, continuous learning, automation, and mentorship.

Overall, the models aim to guide professionals in the statistical programming function in navigating the complexities of Pharma 4.0, fostering individual development and effective team collaboration to meet the evolving demands of the pharmaceutical industry in the fourth industrial revolution. To put things in perspective, pharma industry is already making a strong headway into the new world. We have come a long way from general medicine to personalized medicine by using modern genomics techniques like CRISPR, mRNA silencing, biologics etc. Regulatory bodies are making a secure move towards Artificial Intelligence in Pharma. At statistical programming point of view, we can see major differences in data and standards for the moment. However, with further advancements in drug discovery, we will see a massive shift in clinical research, how we work, what matters in healthcare and timelines. All of which will have a direct impact on the statistical programming function. Therefore, there is a need to reflect on where we are, what we have done so far, what do we need to focus on, and how we are going to align with the fourth and all upcoming industry revolutions, while supporting development in individual and team's interests.

REFERENCES

1. What is the fourth industrial revolution? | World Economic Forum (weforum.org).
2. http://dx.doi.org/10.15265/IY-2016-052.
3. Pharma 4.0: Riding the Fourth Industrial Revolution Wave (linkedin.com)

4. Popov VV, Kudryavtseva EV, Kumar Katiyar N, Shishkin A, Stepanov SI, Goel S. Industry 4.0 and Digitalisation in Healthcare. Materials (Basel). 2022 Mar 14;15(6):2140. doi: 10.3390/ma15062140. PMID: 35329592; PMCID: PMC8953130.
5. Servais L, Camino E, Clement A, McDonald CM, Lukawy J, Lowes LP, Eggenspieler D, Cerreta F, Strijbos P. First Regulatory Qualification of a Novel Digital Endpoint in Duchenne Muscular Dystrophy: A Multi-Stakeholder Perspective on the Impact for Patients and for Drug Development in Neuromuscular Diseases. Digit Biomark. 2021 Aug 5;5(2):183-190. doi: 10.1159/000517411. PMID: 34723071; PMCID: PMC8460979.
6. FDA. Digital health technologies for remote data acquisition in clinical investigations. www.fda.gov/media/155022/download (2021).

ACKNOWLEDGMENTS

I would like to thank Prabhu Kuppuraj and Krishnendu Biswas for providing me with the opportunity to work on this paper. Thank you to my colleagues for enabling me to comprehend the specifics of the statistical programming function, which gave rise to the concept for this paper.

RECOMMENDED READING

Shaping the Future of the Fourth Industrial Revolution – Klaus Schwab

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Ritika Aggarwal
Novartis Healthcare Pvt. Ltd.
Salarpuria-Sattva Knowledge City, Raidurg, Madhapur
Hyderabad / 500081
Phone: +91 7065903685
Email: ritika.aggarwal@novartis.com
LinkedIn: Ritika Aggarwal