International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Modernizing Clinical Data Management with Artificial Intelligence

Aditya Gadiko

Saama Technologies, 900 East Hamilton Avenue, Campbell, 95008, CA, USA
Email: *Aditya.Gadiko[at]aama.com*

Abstract: ***The recent decade has witnessed a threefold es- calation in the volume of clinical trial data, posing significant challenges for sponsors and CROs tasked with its management. With a substantial proportion of trial data now originating from external sources beyond traditional EDC systems, there is a pressing need for sophisticated data handling strategies. Additionally, data inflow velocity has outpaced conventional re- porting methods' capacity to facilitate prompt and informed trial decision-making. This paper contends that Artificial Intelligence (AI) is indispensable for modernizing clinical data management. AI's adeptness in processing expansive datasets equips it to swiftly unearth anomalies, streamline patient recruitment, and enhance the fidelity of data analysis. By automating the labor- intensive aspects of data management, AI liberates research teams to concentrate on interpreting complex data, thereby bolstering the safety and efficacy evaluation of new therapeutics. Adopting AI in clinical data management is not an endeavor free from apprehension, particularly within an industry that prizes regulatory compliance and job security. Nevertheless, AI is posited not as a usurper of human roles but as a facilitator that enriches the acumen of clinical professionals. This advancement encourages a synergistic relationship between technology and expertise, fostering more efficient and robust clinical trials. In the milieu of escalating data demands, AI is an essential tool for clinical researchers, enabling them to navigate the complexities of modern trial data and expedite the journey of new drugs from conception to market.***

Keywords: Clinical Trials, Data Management, Drug Discovery, Artificial Intelligence, Data Automation, Data Review

1. Introduction

The last decade has marked a seismic shift in the landscape of clinical trials, characterized by an unprecedented tripling of data volume [1]. This surge has placed considerable pressure on sponsors and Contract Research Organizations (CROs) to efficiently collect, clean, and interpret sprawling datasets, which are increasingly sourced from diverse systems external to traditional Electronic Data Capture (EDC) systems [2]. The ramifications of this data expansion are twofold: there is an immediate demand for innovative methodologies capable of managing the dispersion of data, and a need to expedite the pace at which this data is processed to inform timely trial decisions. The prevailing manual data management techniques, developed in an era dominated by EDC and predicated on its anticipated role as the primary data repository, are now being challenged. These processes, designed for less voluminous data and slower accrual rates, are proving inadequate against the backdrop of today's high-velocity, diverse clinical trial data. The result is a burgeoning inefficiency that threatens the integrity of trials and the speed with which new treatments can be brought to market.

Enter Artificial Intelligence (AI)—the harbinger of a new era in clinical data management. AI stands as a transformative technology that addresses the dual challenge of volume and velocity by enabling rapid, scalable, and sophisticated data analysis [3]. This integration of AI into the fabric of clinical trials represents a pivotal shift towards more dynamic, real- time data management practices. In redefining the role of AI within clinical trials, this paper will explore how AI technologies not only augment existing data management practices but also serve to streamline the roles of medical reviewers and data managers, ensuring that clinical trials are conducted with greater accuracy and efficiency, all while adhering to the stringent regulatory standards of the industry.

Current Challenges in Clinical Data Management

The landscape of clinical data management is undergoing a significant transformation. In the span of the last decade, the volume of clinical trial data has burgeoned by a factor of three [4], placing an immense burden on sponsors and Contract Research Organizations (CROs) to not only collect and aggregate this data but also to ensure its cleanliness and analytical clarity. Complicating this scenario is the fact that over 70% of clinical trial data now emanates from sources beyond the traditional confines of Electronic Data Capture (EDC) systems, indicating a dispersion of data across a myriad of platforms.

The alacrity with which data is now amassed has also seen a notable acceleration, rendering the previously reliable methods of weekly reporting and batch data analyses insufficient for the exigencies of contemporary trial management. Current manual methodologies, employed to cleanse, reconcile, and interrogate data at such volumes and from such diverse origins, are proving untenable. This inadequacy of resources to carry out these essential tasks is precipitating errors, quality issues, and con- sequent delays in trial timelines. This scenario is compounded by the entrenched data management practices which have developed over the past 30 years of EDC utilisation. These practices, largely reactive and premised on the expectation of EDC's predominance as a data repository, are increasingly incongruous with the growing volume and variety of clinical trial data, including that from labs, sensors, wearables, and Electronic Health Records (EHR).

The industry must grapple with the potential plateauing of EDC's usefulness, given its limitations in processing the

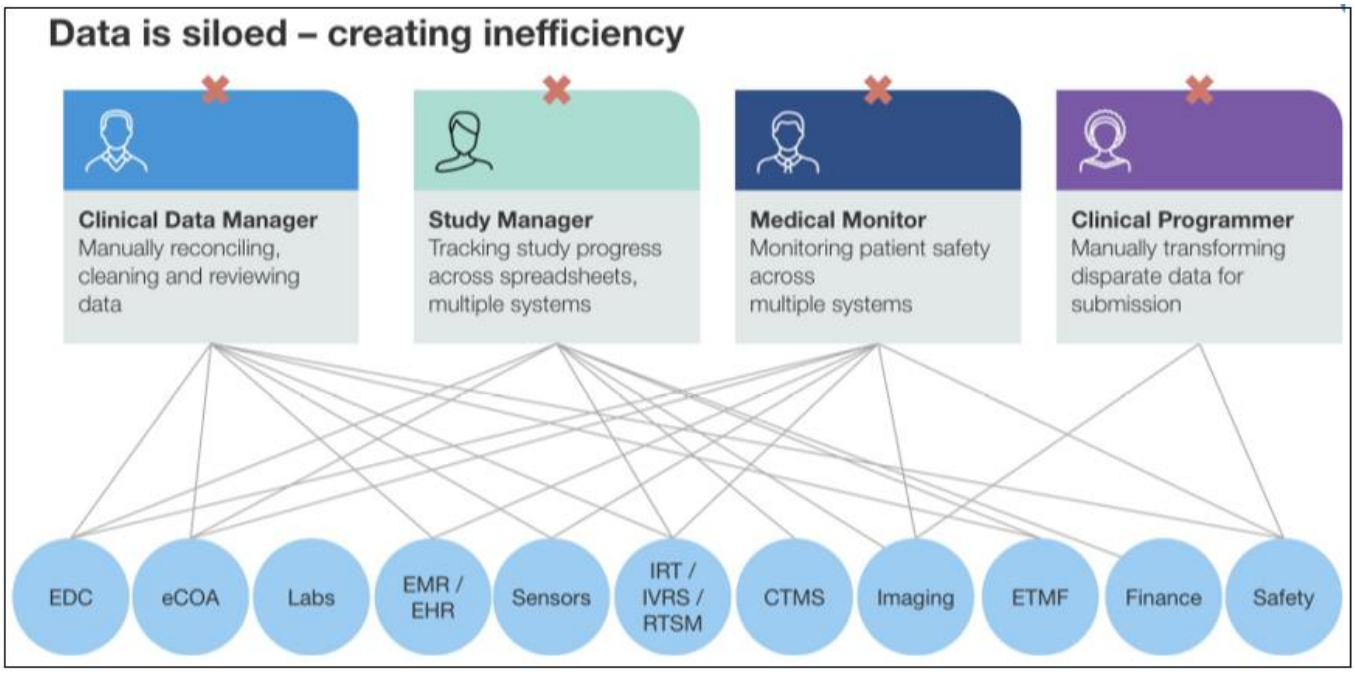

Figure 1: Disparate data sources creating inefficiencies in workflows

ever-expanding cache of non-case report form (CRF) data [5]. This necessitates the advent of novel solutions attuned to the demands of future studies. The inertia of traditional manual data review and EDC query processes, though not inherently flawed, is becoming obsolete—not for lack of efficacy but due to a deficit in speed and efficiency required for modern clinical trials. Amidst the prevailing industry trepidation towards change, particularly in a stringently regulated environment, there is a parallel dialogue regarding the implications of automation, especially regarding job security in the wake of artificial intelligence and machine learning innovations [6]. The focus should pivot from apprehensions of technological redundancy to the strategic adoption of these technologies, enhancing job performance and expediting drug development. The reality is stark: without these technological aids, the sheer scale of clinical trial data remains unmanageable. As we continue to delve into this transformative era, it is imperative to emphasize that legacy processes can no longer sustain the surging volumes of data and that the advancement of clinical trials will increasingly depend on the strategic integration of automation and artificial intelligence into data management practices.

Understanding Clinical Data Automation

Clinical data automation is the linchpin in the cogwheel of contemporary clinical data science, facilitating a paradigm where data volume and complexity are escalating at an unprecedented rate [7]. It is a crucial component in ensuring that the burgeoning datasets in clinical research can be harnessed effectively at the analytical stage. The trajectory of data science within the clinical research sphere is unequivocally towards a future where vast data streams are not just collected, but also meticulously monitored and purified through innovative methodologies.

Automation stands as an indispensable element, ensuring that the influx of data into clinical trials is not only manageable but also actionable. Examples of clinical data automation in practice include the aggregation and mapping of data from a plethora of sources, the identification of duplicates or gaps in data, and the discernment of trends, patterns, and anomalies within high-dimensional datasets. This is complemented by the ability to trace potential adverse events and formulate queries accordingly.

When leveraged appropriately, clinical data automation endows research teams with the ability to derive substantive insights from vast data reservoirs, facilitating more enlightened and expedited decision-making. Analytic technologies play a vital role in structuring and visualizing data, thereby enabling teams to swiftly recognize and react to emergent trends or issues within their studies.

Integral to this process are tools like artificial intelligence and machine learning, which excel at parsing through extensive datasets expeditiously to distill the pertinent from the superfluous, including proffering recommendations or prognoses for consideration by human researchers [8]. This technological aid in data processing is not about usurping the role of human analysts but enhancing their capabilities to navigate the complexities of modern-day clinical trials efficiently.

The Role of Human Insight in Clinical Data Automation

Clinical data automation is designed to augment the decision-making process of researchers, guiding them towards optimal outcomes grounded in robust data analysis. The quintessence of automation in this context is supportive collaboration rather than substitution. The nuanced judgment and critical reasoning of human researchers remain paramount, with technology serving as a tool to enhance, not eclipse, human expertise [9]. Artificial intelligence (AI) and machine learning (ML) are formidable in their computational ability, yet they function within the parameters set by human input. The quality of the output is thus directly tethered to the quality of the input; an echo of the age-old adage, "garbage in, garbage out." These advanced tools mirror the biases and inaccuracies their

human operators imprinted upon them during their training phase [10]. Therefore, the promise of AI and ML in clinical research is contingent upon meticulous and unbiased data provision.

Moreover, the regulatory rigor that typifies the clinical research industry inherently precludes the delegation of critical decision-making to automation. The idea of endowing soft- ware with autonomous authority to navigate through data and enact decisions on behalf of researchers is, at present and for the foreseeable future, beyond the realm of feasibility. This is not a mere technological limitation but a deliberate choice to ensure that clinical research remains firmly under the stewardship of human wisdom, with automation as an aid, not an arbiter.

Strategies for Effective Automation Deployment

The domain of clinical data management is currently undergoing a profound transformation, impelled by the continuous accrual of data from varied digital health technologies like ePRO, eConsent, telehealth, and wearable devices. The sheer volume of data generated demands an inordinate amount of time for manual review—a process becoming increasingly unsustainable. Consider that an average manual data review and query writing could consume 30 minutes per query. Extrapolated over 2000-3000 queries per study, this task could extend to an exorbitant 50-60 days to completion.

Figure 2: Automation vs Manual Processes

Artificial Intelligence (AI) models, trained to scrutinize data corresponding to standard clinical measures, propose a solution to substantially ease this burden. Such technological facilitation could truncate the extensive manual review period down to a mere fortnight or less. This operational efficiency grants research teams the latitude to delve deeper into the data, discerning trends and patterns that automated processes bring to light, thus expediently identifying potential issues or ad- verse events. Furthermore, AI aids in preemptively addressing factors affecting patient recruitment and retention, enhancing the potential to forecast patient withdrawal likelihoods.

In this new paradigm, data managers, and the research team at large, are prompted to reassess their operational roles. Organizations are realigning the role of data managers to focus more on data quality than routine tasks amenable to automation. The modern data manager is becoming a direct actioner of data insights rather than a passive conveyor of problems. This paradigm shift requires a holistic approach to data management—one that embraces comprehensive data acquisition strategies and monitoring. Analogous to aerial forest surveillance identifying areas susceptible to wildfire, a high-level overview facilitated by AI allows for a more strategic assessment of the trial's health than a narrow focus on individual data points could.

However, this transition to AI-driven automation is not without its challenges. It requires a fundamental reconfiguration of existing processes to accommodate the integration of AI tools. Effective clinical data automation initiatives necessitate partnerships with experts who understand the intricacies of workflow modification and offer robust support and training. This collaborative approach ensures that automated functions are both purposeful and calibrated to extract the maximal value from clinical data. Moreover, as data complexity escalates, proficiency in programming languages becomes invaluable, allowing teams to interface with AI models and tailor data transformations directly. Finally, adopting system-agnostic tools becomes crucial, ensuring compatibility with data from diverse sources and fostering seamless integration into the study's analytical framework.

2. Conclusion

In conclusion, the profusion of data in contemporary clinical trials, driven by innovative sources such as eCOA, wearable devices, and sensors, extends beyond the traditional Electronic Data Capture (EDC) systems, substantially enhancing our capabilities for data collection and analysis. The imperative for clinical researchers in this data-rich environment is to adeptly harness technology, particularly automation, to manage the deluge of data from varied and proliferating sources. This technological embrace should not be viewed as a harbinger of obsolescence for researchers but rather as an empowerment tool, augmenting their ability to perform their roles with heightened efficacy. Crucial to this paradigm is a discerning approach to automation—knowing what to automate for maximum efficiency while ensuring that the automated processes are seamlessly integrated into the organizational fabric. It is about striking a balance between technology and human expertise, ensuring that automation enhances rather than supplants the human element. Effective partnerships with seasoned experts in the field are vital for successfully integrating clinical data automation. These collaborations are instrumental in establishing a robust framework for automation that aligns with organizational workflows, ensuring that the full potential of the amassed data is realized. As the horizon of clinical research expands, it becomes clear that the future lies in the strategic application of automation and AI, not as replacements for human insight but as sophisticated tools that reify the ingenuity of researchers. Therein lies the pathway to extracting the utmost value from our collective data and advancing the frontier of clinical discovery.

References

[1] 2022. [Online]. Available: https://www.appliedclinicaltrialsonline.com/ view/the-future-of-clinical-trials-turning-data-chaos-into-trial-intellige nce
[2] Z. Lu and J. Su, "Clinical data management: Current status, challenges, and future directions from industry

perspectives," *Open Access Journal of Clinical Trials*, vol. 2, pp. 93–105, 2010.
[3] Gazali, S. Kaur, and I. Singh, "Artificial intelligence based clinical data management systems: A review," *Informatics in Medicine Unlocked*, vol. 9, pp. 219–229, 2017.
[4] 2022. [Online]. Available: https://www.globenewswire.com/news-release/2021/1/12/2157143/0/en/Rising-Protocol-Design-Complexity-Is-Driving-Rapid-Growth-in-Clinical-Trial-Data-Volume-According-to-Tufts-Center-for-the-Study-of-Drug-Development.html
[5] M. N. Zozus, *Journal of the Society for Clinical Data Management*, vol. 1, no. 1, pp. 1–22, 2021.
[6] 2022. [Online] Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN
[7] pp. 29–29, 2022. [Online]. Available: https://scdm.org/wp-content/uplo ads/2022/06/SCDM-Automation-of-CDM-Driven-Activities-FINAL.p df
[8] 2022. [Online]. Available: https://www.saama.com/ai-delivers-roi-clinic al-trials-data-quality-compliance/
[9] M. H. Jarrahi, "Artificial intelligence and the future of work: Human- AI symbiosis in organizational decision making," *Business Horizons*, vol. 61, pp. 577–586, 2018.
[10] R. Schwartz, A. Vassilev, K. Greene, L. Perine, A. Burt, and P. Hall, "Towards a Standard for Identifying and Managing Bias in Artificial Intelligence," *NIST Special Publication*, vol. 1270, pp. 1–86, 2022.