The Power of Automation in Clinical Data Validation

Accelerating Regulatory Submission Timelines and Improving Submission Quality

Yoran Bar, CEO and Co-Founder of Beaconcure

Orit Veinerman, Vice president of Marketing, Beaconcure

Published: January 2021

Introduction: Applying Artificial Intelligence and Machine Learning to Regulatory Submissions

The use of artificial intelligence (AI) and machine learning (ML) in the life sciences is growing at an extraordinary rate. In the pharmaceutical industry, there is interest in implementing AI-driven solutions to discover new drugs and increase the speed with which they are brought to market[1].

"AI holds enormous promise for the future of medicine. We're actively developing a new regulatory framework to promote innovation in this space and support the use of AI-based technologies." -- Dr Scott Gottlieb, the former FDA Commissioner, April 2018[1]

According to PAREXEL, a leading innovator of global biopharmaceutical services, the amount of data collected in 400 trials and used for regulatory submissions is about 160 terabytes. However, with the influx of wearable devices in clinical trials, a single study trial may reach volumes 12 to 20 times higher[3].

That means more data is being generated and must be collected, monitored, managed and shared. Using traditional methodologies, clinical trial timelines have lengthened, and costs have risen dramatically in recent years. Those involved in managing trials - from sponsor companies to contract research organizations (CRO) and academic institutions - must continuously strive to reduce complexity, streamline business processes and workflows, and increase efficiency across the clinical trial lifecycle[4]

With the growing quantities and complexities of trial data, validation at scale and on time for regulatory submission is becoming increasingly challenging and costly. Automation of routine processes, which saves time and eliminates errors, is the most evident path to a solution. This is especially true as automated AI/ML-based technology rapidly develops, pushing the boundaries of what can be achieved from bench to bedside. It is facilitating entirely new approaches to data generation, management, and analysis of enormous data sets, empowering clinical teams to make faster and more reliable decisions.

The Challenge: Long, Tedious and Manual Processes

Time and quality are critical in regulatory approvals. In fact, the FDA is shifting from stressing compliance above all to prioritizing quality.

Preparing a submission for a vaccine or drug takes about five to eight months. All written documents and the associated tables, figures and listings must support the conclusions presented regarding dosage, efficacy and safety. These clinical claims and data will all be examined by a zero-tolerance regulator. To address regulatory requirements, the industry allocates many resources and large budgets.

What makes that aspect of submissions so complex and challenging?

Lack of Standardization

Delivering high-quality validated clinical data to the regulators can become a major concern when the sources are not standardized. Data may originate from different sources, in different formats, be presented in a variety of table layouts, and be dependent on manual classification by multiple programmers. Another common scenario is, for example, when a pharmaceutical company acquires another company and must integrate clinical studies developed by separate teams and third parties to prepare for regulatory submissions.

Stressful Timelines and Workloads

Successful delivery of analysis outputs for a clinical trial depends on a strong working relationship among members of a biostatistics team, which typically includes biostatisticians and programmers. As clinical trials and their analyses get more complex, these teams face increased pressure to produce outputs efficiently and quickly. It is at that stage that communication between the statisticians and programmers tends to break down, precisely when it is needed most. This results in stressful last-minute work and rework[5].

Due to the intense focus researchers place on setting up trials and making sure they run smoothly and safely, analysis of all the collected data may be inadvertently short-changed. If the trial takes longer than expected due to low enrollment or other issues, there is often a pressure to shorten the period or extent of data analysis. Commercial entities may also push to finalize the clinical analysis stage as early as possible, as every passing day risks affecting the price of the final product and reducing its market share. In such scenarios, analysts or organizations sometimes take unnecessary risks in order to deliver faster results for regulatory submissions.

Risk Management

Quality control, especially data validation, is fundamental to ensuring both correct results and sound interpretations of clinical studies. Even in the current environment of lean workforces, increased workload, high demand for statistical programmers, tighter budgets and shorter timelines, high-quality output is still expected. Given these challenges, organizations are forced to find ways to improve efficiency without compromising on quality.

The current gold standard for validation among many companies in the data analysis industry is 100% double programming of all analysis datasets, tables, listings and figures. This approach, however, may no longer be necessary, does not resolve cross-table discrepancies, typically does not

account for unanticipated data changes during clinical trials, and, we would argue, is not the best method of obtaining the highest quality results[4].

In an ideal world, every data point collected, and every program written for a clinical study report would be validated 100% by some sort of independent review or reproduction. It is our experience, however, that even when a full independent reproduction of a report has been performed, errors can still arise from data anomalies, spelling errors, formatting errors, incorrect interpretation of specifications, etc. Furthermore, identification and resolution of these remaining errors can take an increasingly long time. Anyone who has participated in the process knows that it is time-consuming and seems never-ending[4].

Lack of Transparency

In typical validation processes, there is no transparency regarding how many data cross checks have been performed or how long the process takes. This poses the risk of additional discrepancies when biostatisticians end up working on competing versions of tables or duplicating their efforts.

The Solution: Automation Supported by ML

How might the major clinical data validation challenges be relieved? The most comprehensive and effective solution is automation supported by ML. AI or ML technologies have already been proven across many industries, and they are considered a game-changer in the world of drug discovery and development. Applied to data validation, ML and automation can be fundamental to rapidly ensuring sound interpretation of clinical data and, therefore, more efficient regulatory submissions.

Verify is Beaconcure's automated validation solution, supported by ML and created specifically for the life science industry. It converts various clinical data formats in any layout into a semantic and dynamic database, to which any required segmentation rule, crosscheck and analysis can be applied. All defined errors and anomalies are identified, with 99.7% accuracy, **in a matter of hours**, drastically reducing clinical data processing timelines and increasing the quality of the output.

How does the technology work?

Verify validates statistical outputs (SAS®) automatically by applying various algorithms to the processed data. The verification algorithms use the base table processing information to identify groups and sub-groups in the data, with the capability of validating single and cross-table data as needed. The system can then flag discrepancies and direct the user to the relevant table for follow-up action and resolution of identified discrepancies.

Unlike other AI solutions, Verify's implementation process only requires the customer to provide a list of required data checks once, when the system is being customized to meet their needs. This unique approach makes using Verify fast, effective and easy.

Formats

Verify can extract, identify and manage data from tables and text in any of the following formats: TXT, RPT, HTML, RTF, PDF, Docx.

The Power of Automation: Saving Time, Improving Quality

Using Verify in the validation process significantly improves data accuracy and quality, while allowing organizations to meet tight regulatory submission deadlines.

- 99% of data discrepancies will be flagged.
- 100% of data duplication will be flagged.
- 100% of data failures will be flagged.

Verify's ML technology supports automation in any therapeutic area, eliminates time-consuming, manual cross checks and mitigates risk before it becomes an issue. Using ML technology allows Verify to learn from the users' feedback and improve the algorithms almost in real-time. Users can also track extracted and validate data, identify which checks were applied and what the results were. In the event of a question during regulatory review, the organization can reply quickly and easily.

Moving from a manual process to an automated one with Verify accelerates the regulatory submission and approval timeline, improves data integrity, frees up resources and reduces costs.

Conclusion

The reason for validating SAS® results is simple: important decisions are made based on analyses generated by SAS programs

Using automated validation supported by ML to replace manual processes goes a long way toward addressing the pressures arising from demanding timelines and heavy workloads.

The key to success in implementing automation is the combination of the human factor and artificial intelligence or machine learning. More and more pharmaceutical companies and CROs are using AI or ML technologies to streamline and improve processes. By adopting an automated platform, companies can increase the efficiency of their clinical trial data validation and submission, avoiding regulatory delays and expediting the availability of life-saving therapies.

For more information on implementing a state-of-the-art validation solution, visit Beaconcure today https://beaconcure.com/ hello@beaconcure.com

1. The FDA turns to AI to accelerate drug development and approval. By Dr Nicola Davies, April 23, 2020. The FDA turns to AI to accelerate drug development and approval
2. Digital Automation in Clinical Trials: The Promise and Potential. September 29, 2015 Digital Automation in Clinical Trials: The Promise And Potential | American Pharmaceutical Review - The Review of American Pharmaceutical Business & Technology
3. Billions Will Be Poured into AI Drug Development. By Yiannis Mouratidis, Nov 14, 2018, Forbes. Billions Will Be Poured into AI Drug Development
4. Risk-based Validation in Clinical Trial Reporting: Focus on What Matters Most. By Amber Randall and Bill Coar. PharmaSUG 2018 - Paper SI-02. Risk management.pdf
5. Improving the Relationship between Statisticians and Programmers in Clinical Trial Studies. By Mai Ngo, Mary Grovesteen, PharmaSUG 2019 https://www.pharmasug.org/proceedings/2019/LD/PharmaSUG-2019-LD-207.pdf